

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2002, 13:49:49 ; Search time 1879.78 Seconds  
(without alignments)  
13459.119 Million cell updates/sec

Title: US-09-724-797-35

Perfect score: 1209

Sequence: 1 gtgtgtgtgtgtgtgtgtgtgtgtac.....cgttcacgcgcgcgtgtga 1209

Scoring table:

IDENTITY-NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Genem1:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_pro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pal:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No. Query Match length DB ID Description

C	1	250	20.7	39524	1	SCH10	AL049754 Streptomyces
C	2	245	20.3	3078	1	SERCP450A	M83110 Saccharopol
C	3	220	18.2	10060	1	AB004675	Pseudomon
C	4	220	18.2	24268	1	AB070947	Streptomy
C	5	218.6	17.8	44274	1	SCH63	AL442628 Streptomy
C	6	215.6	17.8	190050	1	AL646080	Streptomy
C	7	214.4	17.7	63734	1	AF127374	Streptomy
C	8	212.2	17.6	6085	6	AB0304	Streptomy
C	9	212.2	17.6	6085	6	AR144763	Sequence 4
C	10	204.4	16.9	1257	6	AR097457	Sequence
C	11	204.4	16.9	14759	6	AB004755	Pseudomon
C	12	203.6	16.8	2795	6	AB018074	Streptomy
C	13	203.6	16.8	2795	6	ES9713	Process for
C	14	198.2	16.4	36028	1	SCE41	AL442120 Streptomy
C	15	195.6	16.2	1400	1	STMOLEP	L37200 Streptomy
C	16	195.6	16.2	50937	6	AR159871	Sequence
C	17	194	16.0	2243	1	SEREXFEGH	AR159871 Sequence
C	18	193.8	16.0	1700	1	SGSOYBC	M54983 Saccharopol
C	19	193.8	16.0	1735	6	I15434	X63601 S.griseus s
C	20	193.2	16.0	2177	1	AF102510	I15434 Sequence 10
C	21	192.4	15.9	2218	1	SRE250199	Streptomy
C	22	192.4	15.9	9389	1	SRE250581	Streptomy
C	23	190.8	15.8	10167	1	STU08223	Streptomy
C	24	187.2	15.5	104326	1	AB070940	Streptomy
C	25	183.8	15.2	4254	1	SEERYABIO	X60379 S.erythraea
C	26	182.8	15.1	6690	1	AF322179	Streptomy
C	27	179.2	14.8	16057	1	AE007165	Streptomy
C	28	179.2	14.8	47852	1	MTV023	Streptomy
C	29	179	14.8	1470	1	AF087022	Streptomy
C	30	179	14.8	1688	1	SYM50BCB	M32239 S.griseolus
C	31	177.4	14.7	4342	1	AF079139	Streptomy
C	32	177	14.6	55972	1	AF386507	Streptomy
C	33	173.8	14.4	145911	1	AP003014	Mesophilzo
C	34	171.8	14.2	8366	1	AF072709	Streptomy
C	35	171.4	14.2	38340	1	SC4C2	Streptomy
C	36	171	14.1	25797	1	SC2G58	Streptomy
C	37	170.6	14.1	78210	1	AB070949	Streptomy
C	38	170.2	14.1	5774	1	AB070948	Streptomy
C	39	169.6	14.0	107379	1	SHGCP1R	Streptomy
C	40	168.6	13.9	17228	1	SC3F60	Streptomy
C	41	166.2	13.7	46166	1	SC8F11	Streptomy
C	42	164.8	13.6	25617	1	AF170880	Streptomy
C	43	163.6	13.5	38426	1	SCF55	Streptomy
C	44	160.8	13.3	63734	1	AF127374	Streptomy
C	45	160.4	13.3	8487	1	AB048795	Streptomy

#### ALIGNMENTS

RESULT 1  
LOCUS SCH10 39524 bp DNA linear BCT 04-MAY-1999  
DEFINITION Streptomyces coelicolor cosmid H10.  
ACCESSION AL049754  
VERSION AL049754.1 GI:4753846  
KEYWORDS ABC transporter; adenylosuccinate synthetase; aspartate aminotransferase; ATP-dependent protease ATP-binding subunit; cytochrome P-450 hydroxylase; fba; fructose 1,6-bisphosphate aldolase; lipase/esterase; oxidoreductase; purA; RNA polymerase sigma factor; spermidine synthase; transmembrane efflux protein; two-component regulator; two-component sensor kinase; uridine 5'-monophosphate synthase.

SOURCE Streptomyces coelicolor A3(2).

ORGANISM Streptomyces coelicolor A3(2)

REFERENCE Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

Redenbach, M., Kieser, H.M., Denapate, D., Eichner, A., Cullum, J., Kinash, H., and Hopwood, D.A.

A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome  
Mol. Microbiol. 21 (1), 77-96 (1996)

JOURNAL



```
/gene="SCH10..06c", hypothetical protein, len: 163 aa;
/note="SCH10..06c, hypothetical protein, len: 163 aa;
unknown function, probable CDS suggested by positional
base preference, gc frame analysis and amino acid
composition"
/codon_start=1
/transl_table=11
/label="SCH10..06c
/product="hypothetical protein"
/protein_id="CAB42015.1"
/db_xref="GI:4733852"
/translation="MRKPTRTIMLAAVAVVYVGAFFYVDGHRVSAHPAGDKDE
IVSRIPDIDPGASDMALGDVASMGGKAFVPCGAKELDPNNLVCYTAGVWVIDE
ARLKRDSVYLRTFYGRSPAVEFYSGPREVEGLILVADPAVMWIPQERKICLDAA
TVL"
complement(3412..4695)
/gene="SCH10..07c"
/note="purA"
/complement(3412..4695)
/gene="SCH10..07c"
/note="SCH10..07c, purA, probable adenylosuccinate
synthetase, len: 426 aa; similar to many e.g.
SW_PUR_A_ECOLI (EMBL:J04199), purA, Escherichia coli
adenylosuccinate synthetase (431 aa), fasta scores: opt:
1421 z-score: 1608.0 E(): 0, 50.5% identity in 420 aa
overlap. Contains Pfam match to entry PF00709
Adenylosucc_synth, Adenylosuccinate synthetase, score
742.40, E-value 1.9e-219, PS00513 Adenylosuccinate
synthetase active site and PS01266 Adenylosuccinate
synthetase GTP-binding site"
/codon_start=1
/transl_table=11
/label="purA"

Query Match      20 7%; Score 250; DB 1; Length 39524;
Best Local Similarity 53.9%; Pred. No. 2.8e-12;
Matches 620; Conservative 0; Mismatches 500; Indels 30; Gaps 4;

QY 14 ccgtgaccgctgcatccaccagcgcgcgtgacccctaccctccacc 73
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11617 CCGGTCGGCGCTTGACCCCGTGGAGACCCCGCTTCGTCGCCGACCCCTTCCG 11558

QY 74 actggtctgtgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 130
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11557 CCGAGCTGCGGGGCCGGGGGCGCTCTACTACGAGCGAGGACCACTGCGTGTCC 11498

QY 131 ccgccttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 190
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11497 CGCACACGCGGAGCTCTCCGCGCTGCTGCGACCGCGCTTCGGCGGACCTACCGAGC 11438

QY 191 gccggtcttgaccgacactgctgcgcgcgcgcgcgcgcgcgcgcgc 230
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11437 ACCGTTTCACGACGACGAGACTTGGCGCGGACGCCGCCGCCGCGGAGCAGACCTTCC 11378

QY 231 -cgagatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 289
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11377 ACACGCTTACGACGACGAGGATGCTGACCTGAGCGCGCGGACCAACACCGTATCCGGC 11318

QY 290 ggggtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 349
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11317 GCGTGGTGTGGAAGGCTTTCACCGCGCGGACCGTGGAGCGGCTGAAACCGTACGTGCACG 11258

QY 350 gcaacgttgagacgactgtctgcgcgcgcgcgcgcgcgcgcgcgcgc 409
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11257 GCCGTGGGAGGACACTGTGTGGCGGTGTGGCGGCGGCGGCGGATCTCTCACCG 11198

QY 410 agctgagctaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 469
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11197 ACGTGGCGGAGCGCGCTGCGCGGTGATCGCCGAGAGATGCGGACATCCCGAGGTCCG 11138

QY 470 actggggggcgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 529
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11137 ACCGGGCGCGCGTGGCGCCCTGTGTCGGCGGAGATCTCGGGAGTCTAGACCTAACCCGT 11078
```

```
QY 530 ggcgcgagagacatgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 589
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11077 CCGAGGAGAGCGGGCGGCGAAGCGGTCGGGCGCTCGCGACTTCTCGACTACTCTCGCG 11018

QY 590 gggcgccctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 649
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11017 CGCTGATGCGGGCGCGCGCGGCGAAGAGCGCGGCGACACACTGATCTCGGTCTGATCGCG 10958

QY 650 cccacgacgc---cgscctatgatgcgaacacgctcgaacgcgcgcgcgcgc 706
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10957 CCCACGACGAGGAGACGACCGGCTCACCAGAGAGATGATCTCGACTCGCTCTCTCC 10898

QY 707 tcttcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 766
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10897 TGAACGCGGCTCACGAGGCGCACGCTGACGCCACCAACAGCGCTGGCGCTCTTCC 10838

QY 767 cgcacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 826
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10837 GCCACCCCGACGACTGCGCGCGCGCGCGCGCGCACACTCTCTGTCGCGCGCGTCCG 10778

QY 827 aggaagtgcctgcgtacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 886
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10777 AGGAGCTGATGCGCTACGACACCGCGCTCAGCTTTTCAACGTTGGTCTCGAGACAGA 10718

QY 887 tggagctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 946
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10717 TCGAGATCGACGCGCACGACCGCTCCCGCGCGCGGAGGTGGAGTGTTCGGCTCCG 10658

QY 947 cgcacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1006
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10657 CCAACACGACCGCGGCGGCTTTCACGACCGGAGACCGCTCACTCACCCCGCGGACA 10598

QY 1007 tcccgctgatgtctctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1066
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10597 ACCCGACATCTCTTCAAGCGCGCGGCGCATCACTGATGATGCGCGCGCGCGCGCGGA 10538

QY 1067 cgaagctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1123
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10537 TCGAAGTCCGGCGGCTCAATGAGTCTCTGTAAGCGGCGCGCGCTGCGCTGGCGG 10478

QY 1124 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1133
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10477 CGGACCGGGA 10468

RESULT 2
SERCP450A 3078 bp DNA linear BCT 26-APR-1993
LOCUS Saccharopolyspora erythraea ORF 1 gene, partial cds: cytochrome
DEFINITION P-450 gene, complete cds; ORF 2 gene, partial cds.
ACCESSION M83110.1 GI:152682
VERSION M83110.1 GI:152682
KEYWORDS cytochrome P450.
SOURCE Saccharopolyspora erythraea
ORGANISM Saccharopolyspora erythraea
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Pseudonocardiales; Pseudonocardiales;
Saccharopolyspora.
REFERENCE 1 (bases 1 to 3078)
AUTHORS Andersen J.F. and Hutchinson C.R.
TITLE Characterization of Saccharopolyspora erythraea cytochrome P-450
genes and enzymes, including 6-deoxyerythronolide B hydroxylase
JOURNAL J. Bacteriol. 174, 725-735 (1992)
MEDLINE 92121109
FEATURES
source Location/Qualifiers
1..3078
/organism="Saccharopolyspora erythraea"
/db_xref="taxon:1836"
/tissue_lib="NRRL 2338"
<1..859
/note="ORF homologous to Escherichia coli and Salmonella
typhimurium dnaE gene; putative"
```

```
/codon_start=2
/transl_table=11
/protein_id="AAA26482.1"
/db_xref="GI:152683"
/translation="AGGGRGHRRAOPAGARPVPRVRRGRRRRRVVVAAGAPAVHAG
RYAQAADALLRAGDAGPLRLGPPSGRGAARAAVAVGHHRRARRRRRAGGPGDDR
DLGDPAPHOORPPVDDPGPGRROGAVLPVLVDVLRPRRGDRGEGEGRHOR
ARGHHOHLRRRAGGHLGGDRPNORVRROGAGEGDPAGBRAOAHLAGALGHS
PGARRAAGARHPPRAVQRLLRVREBRAGAEGAAARGVLRDCAVSRRAASTLRA
PRPLRDS"
949..2186
/codon_start=1
/transl_table=11
/product="cytochrome P-450"
/protein_id="AAA26483.1"
/db_xref="GI:152684"
/translation="MTGHEVPDLAFEDDAFADDRNRARRREPVQIRTVNGIDAN
LITREDAVQALLDPRIAKDFERTQOILEKLADERRPESPDLPMLNTDPDHT
RLKLIVKAFTRARVEGLRPIEITDILLRLAGRSEVDLIDEPFLPTIVISELM
GVEDSRDPRFMTVYLDGSOPEAOAOSVAMVEYLIELAKRTREGDDLITLLE
AVEDGDRLSEGLIMVFLIYAGHETTVNLGNCVLSLGNPDOLALRNDPSLLPG
AIEHTLRSPVANGTFRHTFAVRGDDVPEGELVVALGAARDEPREPDPRFD
ITRETTGHAFFGHGHTFCYGAALARLEAOIANGRLERPDLMASPDLLRMRSVL
MRGLEXLPVRPA"
complement(2365..>3078)
CDS
  complement(2365..>3078)
  /note="ORF 2"
  /codon_start=1
  /transl_table=11
  /protein_id="AAA26484.1"
  /db_xref="GI:152685"
  /translation="LIAVATGSTLFVLSQYLDLVGRHSAVVAAMALPLIAGSVVGA
LGAARPARIGRACIVTGFVAVTAFVAGFVLAALGPSOPHILFGILGFGFAGPA
ASTALGAVPADRAGMSALNDTHOOLGIAFGLVAGLISLAVRAPLPTGVPHDAST
SLAATLSFADERTSALADARLAFTLQASATMTMGLACALGAAVAMLILSRGASA
VALTASATARRNLRAGR"
BASE COUNT      463 a      1060 c      1137 g      418 t
ORIGIN
Query Match      20.3% Score 245; DB 1; Length 3078;
Best Local Similarity 52.5% Pred. No. 1.9e-11;
Matches 623; Conservative 0; Mismatches 540; Indels 24; Gaps 3;
QY      8      tcgagtcgctgacgcgctgcatccaccacgacgcgctgacgctgacccctacgcgt 67
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      983      TCGACGACGGCTTGGCGCAGAGACCGACAAACGGTACGGCGGAGCGGAGAGCGG 1042
QY      68      cctaccactgctgctgacgacgacgctgacgctgacgctgacgctgacgctgacg 127
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1043      TGCAGCCGATCCGACGCTAAGCGGCTCGACCGCTGCTGATCACCCTGACGAGGAGC 1102
QY      128      tctccgcctcgcgacgctgacgacgctgctgacgacgacgctcgcgcgcgcgcgcgcgcgc 187
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1103      TCAAGCAGGGCGCTGCTGGACCCCAAGGATCGCCAAAGACTTTCGGGCCACCCAGACATCA 1162
QY      188      tccgcgcttctgacgacacctgctcggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 247
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1163      TCGAAGAAAGAGGCTGGCGACGCGGAGCGAGCGCGGGGCTTCGCCCGACCTCGGCCCGC 1222
QY      248      tcatctgttccagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 307
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1223      ACATGCTCAACACGACCCCGCGGACACACCGGCTGCGCAAGCTGCTGATCAAGCGCT 1282
QY      308      tctcgcgctcgcgctgctgacgacgctgacgacgctgacgacgctgacgacgctgacgacgctg 367
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1283      TCACCGCGCGCGGCTGAGAGGCTTGGCGCGGATCGACGACATCAACCGAGACCTGCTGC 1342
QY      368      tgcgcgcgcgcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 427
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1343      TCGACCGGCTTGGCGCGGCTGCAATGATCTGATCGACGAGTTTCCGCTTCCCGCTGC 1402
QY      428      cgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 487
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
DB      1403      CCATCACGGTATCTCGAGCTGATGGGCTGAGAGACTCGCGCGGACACTTCCGCT 1462
QY      488      gctgctgcgcgacgctggaagccctggaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 547
DB      1463      CGT-----GGACCAACGTCCTGGTGGACGGGTCCGACGCCGAGGCGC 1504
QY      548      gcgcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 607
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1505      AGGCCCAAGGCTCGGCTGGCGATGGTCGAGTACCTGACCGACCTGATCGCCAGAAAGCGCA 1564
QY      608      ggcgtgagggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 667
DB      1565      CCGAGCCCGGTCGACGACCTGTGTGACCCGGCTGCGAGGCGGCTGAGAGCGGCAACCGCC 1624
QY      668      tgaatcgcaacagatcgtcagcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 727
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1625      TCAGGAGGCGCAGCTGATCCGCATGGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1684
QY      728      tggcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 787
DB      1685      CGGTCAACCTGATCGGCAACTGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1744
QY      788      tgcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 847
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1745      CGCTCGCAACGACGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1804
QY      848      cgtcgtgctcagtcacacacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 907
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1805      GCCCGGTGGCGAAGCGGACGTTCCGGCACACCGCGGAGCGGCTCGGCTCGGCAACGTGG 1864
QY      908      tgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 964
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1865      TGAATCCCGAGGAGAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1924
QY      965      gttacgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1024
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1925      GCTTCGAGGACCCCGACCGCTTCGACATCACCAGGAGAACCAACCGGACGCTGCTGCTGCTGCTGCTGCTGCTG 1984
QY      1025      gcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1081
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1985      GCGACGGCATCCACTTTCGCGTGGCGGCGCGCTGCGCGCTGCGGAGCGGACCGACATCGCGC 2044
QY      1082      cgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1141
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      2045      TCGGACAGGCTCCTGAGGCGCTTCCCGGACCTGCGGATGCGCGCTGCGGACGACCTGCTG 2104
QY      1142      cctatcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1188
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      2105      GGTGGCGTTACGCTGCTGATGCGGGGCTGGAAGAACTGCGGCTC 2151
RESULT      3
AE004675/c 10060 bp DNA Linear BCT 30-AUG-2000
LOCUS      Pseudomonas aeruginosa PA01, section 236 of 529 of the complete
DEFINITION genome.
ACCESSION      AE004675 AE004091
VERSION      AE004675.1 GI:948522
KEYWORDS
SOURCE
ORGANISM      Pseudomonas aeruginosa.
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE      1 (bases 1 to 10060)
      Stover,C.K., Pham,X.O., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
      Hickey,M.J., Brinkman,F.S., Huftagle,W.O., Kowalik,D.J., Lagrou,M.,
      Gardner,R.L., Goltry,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y.,
      Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Lardig,K., Lim,R.,
      Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
      Complete genome sequence of Pseudomonas aeruginosa PA01, an
      opportunistic pathogen
JOURNAL      Nature 406 (6799), 959-964 (2000)
MEDLINE      20437337
```



REFERENCE 2 (bases 1 to 10060)  
AUTHORS Stover, C. K., Pham, X.-Q. T., Erwin, A. L., Mizoguchi, S. D., Warren, P.,  
Hickey, M. J., Brinkman, F. S. L., Hutnagle, W. O., Kowalik, D. J.,  
Lagrou, M., Garber, R. L., Goltry, L., Tolentino, E.,  
Westbrock-Wadman, S., Yuan, Y., Brody, L. L., Coulter, S. N.,  
Folger, K. R., Kas, A., Larbig, K., Lim, R. M., Smith, K. A., Spencer, D. H.,  
Mong, G. K.-S., Wu, Z., Paulsen, I. T., Reizer, J., Sailer, M. H.,  
Hancock, R. E. W., Lory, S. and Olson, M. V.  
TITLE Direct Submission  
JOURNAL Submitted (16-MAY-2000) Department of Medicine and Genetics,  
University of Washington Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA  
FEATURES  
source 1. 10060  
/organism="Pseudomonas aeruginosa"  
/strain="PAO1"  
/db\_xref="taxon:287"  
70. .1416  
/gene="PA2472"  
70. .1416  
/gene="PA2472"  
/codon\_start=1  
/transl\_table=11  
/product="probable MFS transporter"  
/protein\_id="AAG05860.1"  
/db\_xref="GI:9948523"  
/translation="MPTPKNNKIDVOAFIDAOVASAFOILRVLSICFLIYAMGPDPA  
AIGFIAPALAHMDSPALSPITIGALAGLACAPAGIADLPFGKSVLLSVLFF  
TUGSALGSLAAKVPALGWESVLLGGGIPLASLPLWACLPESVRFYARAPDAR  
LRRLERLGRLLPDGMAGRELPADESDGTSPLRQIIGAEIRGTLTLLMTFFGLLI  
YLTNMLPLTIGTGSLGEANATISAMFOLGTYLLGALLGSAMDREDAHVLSLAVG  
GALFIIIGLISLHSPFALLACVAGVFCISQOVGNALAADEYPTRSRATGVSMALG  
IGRGSIVGSLGSGALLGLGFGSGLIALLIVIPALLAAVAHVHGRRRRPSFTTL"  
1429. .2073  
/gene="PA2473"  
1429. .2073  
/gene="PA2473"  
/codon\_start=1  
/transl\_table=11  
/product="probable glutathione S-transferase"  
/protein\_id="AAG05861.1"  
/db\_xref="GI:9948524"  
/translation="WQYSEFNSSTYSRVRIALALKGLDYVVVNLROGEOLRPADR  
ORNPWGLPPLVDADGRFSQSLAIIDYDAVQPEPLPLDPLHRAQLLEALIVAC  
DHPNNVRLKTLTGVLGIDAEORWYAHWAEGIAAETLLNHRGAGAFAGAA  
GIVECCIVPOLANARWCDIAPYPALLEGRCLALEAQRASPERQDYPD"  
2115. .3026  
/gene="PA2474"  
2115. .3026  
/gene="PA2474"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAG05862.1"  
/db\_xref="GI:9948525"  
/translation="MTTRIRASLLALPLLPFSFOATALPGMNLGOTSFLDKALPGT  
LFQOVNIYEADPFMDANGRRRLDNRLRLTLATSFHAIISOHLEGAYYGEIIVPV  
LLADPPEBPAGANGILIVSPFLQWQPEFLFRPYWQNLAFOLPFGYDHR  
AINDGSNHSFNHYALTWEFAWGEASARLOYLLPGKRRDAPPLADXYLOPGCAF  
ANYSVSLADAHMRLLGSGSLRQITDERVDGDRDSRERDAIGPALPMFGKSRLL  
ELNAVYHEGARNSEBSRLILRSQYF"  
complement(3008. .4342)  
/gene="PA2475"  
complement(3008. .4342)  
/gene="PA2475"  
/codon\_start=1  
/transl\_table=11  
/product="probable cytochrome P450"  
/protein\_id="AAG05863.1"  
/db\_xref="GI:9948526"  
/translation="MDDAFSEEGSAQPRHDAORPALAPRSDFDIIHTYHPDVADPY  
LTRLIRSAFVPCBDQASIMWISRYADVSACLDRDRSADPALGAGVQGGASNFGR

QQLIPLARFYDNFMFLFNDAFRHTRRLRLEFAPAFGPDVAWRWEARIEVLIELDLSLE  
RRPDLIRDFAPETLITRYAAELFEGPREDETCOLLWGRDLAAGLDAASHGAGQOINR  
SAVAFSDYDORORAGWSDSSPPSGAASIIIDGAMLEAGLGLDIDVAAYAMVMAA  
FETTIMVENATLALITHPDOLILRCEPEALAAVEELIRPDGAVRGVRCOTLEVEE  
IGGORIPREKXWLSFLANRRPEMFARPDRIQORANKQVAFARHGHGTCGLAVLA  
RLELOCALGLVRRFALASEPTDLKWRSSVFRILRLIPIVEGDAKTCG"  
complement(4415. .5185)  
/gene="dsbg"  
/note="PA2476"  
complement(4415. .5185)  
/gene="dsbg"  
/codon\_start=1  
/transl\_table=11  
/product="thiol:disulfide interchange protein dsbg"  
/protein\_id="AAG05864.1"  
/db\_xref="GI:9948527"  
/translation="MRIPNLITIGLITLINAGLATAAEELPAPIRWOKGARIILG  
SPDAPDGLRGVAAEYONOGMALXYLPDCKHYITGTLPAQCKDLSREPLERYAPILA  
KEMQKMEQSANIADGRADAPRVYTLFSDPNCPTCTMFEQARWPDVQVLRHIV  
GITREDSEAKSALIALASKDPQKALHDHROAGASTLKPLAKIPAAVVRQLAGNMELME  
SMGAATPAIFEYLNAGEWMQDQOGAPDOLAEIIGPR"  
complement(5182. .6018)  
/gene="PA2477"  
complement(5182. .6018)  
/gene="PA2477"  
/codon\_start=1  
/transl\_table=11  
/product="probable thiol:disulfide interchange protein"  
/protein\_id="AAG05865.1"  
/db\_xref="GI:9948528"  
/translation="MISFSIGPIVYSLQHLITLALGALLGWLAAAGGRNAPVL  
EULLIGLIVARLAVVRYWYPOYRQDFQOMLDIRGGTLAMPGLAAVLAFLPAMR  
PALRSISGASLGLAEWLSGLGITEKGTLPBELSLRNAAESVQLAFLRPLV  
NLMSWCPCRRKREMPVLQQAENPDVVFLEANGESAETVRHFLQGENLRNLNLF  
DNGQLGOOVGSVALPTVTFYTABGRLLGSHLGELSRSLARYLEAFEPAAAPATSS  
SE"  
complement(6018. .7781)  
/gene="PA2478"  
complement(6018. .7781)  
/gene="PA2478"  
/codon\_start=1  
/transl\_table=11  
/product="probable thiol:disulfide interchange protein"  
/protein\_id="AAG05866.1"  
/db\_xref="GI:9948529"  
/translation="NRVILILMLLPELSQAQPEDDLFAPRGATQCTDFLVEKARFE  
TWERLDSDVOQLRWQIAPGYVLYQKRLREFDLDLPALQPLPFGSHSDSEFGESQVLR  
OSLELTLPAAAGQDLRLGMOCCADAGLCYPPQSOALDLGGTGPAAAGTSEVADQGL  
AGSLQAGNLAWSLLFFGLGLLAFAPCSILPMLDLIAGLVGSGAGPRGILLAGSYV  
LSMALVYAGLGVVALLGNLQAMIQOPMLGSPFALAEFIALPMFGPELQPLPALR  
DRLDGLSRGRKGGSLAGAAALGSLLVGCPMTAPLAGALLYIAOTGNALHGSLVFE  
SLGISTGMPDLIIYVGSRPFKPRGPNMNVKYGVGFLFCTANTILRPLGLGELWIG  
LGALLIVLATRAALHTARGLNRHVLFEAAGCTIGLWGAAMLGAAAGADDPWPRVQV  
YAAANRGATPTASAEHAFPLVSQPELDRQLAARKABQEWVLLDYVADWCVSCRIMK  
OVFAKPDVIALAOGVRLRLDVTADNMASSRLRLHYQVPPPSLIITWIGPEGERARR  
LITGEVDAGFLAHWOATRERG"  
7944. .8624  
/gene="PA2479"  
7944. .8624  
/gene="PA2479"  
/codon\_start=1  
/transl\_table=11  
/product="probable two-component response regulator"  
/protein\_id="AAG05867.1"  
/db\_xref="GI:9948530"  
/translation="MHVLLITEDDDLINSGLVAGLNAAGLTVDRVYSAADTQALLQYAR  
FDVLVLIDGLPDEDELRLQLRQGVDPVLYVTARDAYTDVAGIQAGADYLLKPR  
FDLRLGARLHTLQRRSAGRCVNIIEHGRLSYDSTRETETWLDGPRVLSRREOALLQA  
LLNNRGLTISGEQLKDSYVGGDEVESNALNVHHILRRKLGNALIVQTVAGLGRIGP  
ARGGDDA"  
8621. .9943  
/gene="PA2480"  
8621. .9943

```

/gene="PA2480"
/codon_start=1
/transl_table=1
/product="Probable two-component sensor"
/protein_id="A050868.1"
/db_xref="GI:9948531"
/translation="MSLRRLRLTLLGSAFLLNLSLAAMMLCDLRGMMRSPFORLAA
SARRVAGLILQOLPQPPADQOTTRLSEQLGIPGMACQVSLKEGYLAOSHATPEPA
LDTSASGFQDEINGVLRBSFTVQAGDLRASTDALDERVLPGLSVVAAPVTVAL
IGSLATMLMGAGGSLAPLRKLRIDALTRGGADSTLEPLALPSLRKPLVEFONLFOR
IAQAIERERRRLTGDAHELSPGLTAITTHIQVAMRTGGAANEQALAAEAGADLHRT
LDQILLIARVEGSLSTFDGQVQCSAEDAAVAMHDAGHDVMEFLKADGISEAPLEV
PSYLAIALRLVDNLNRLHTAAGTRVLAETAGCCARFVHDGPIIAETDILPHLTR
RWMHGKSGGSGGLATVQQAIVQCRCORLPDPSRADGLRVTLTEMPLEPR"

```

Query Match	18.2%;	Score 220;	DB 1;	Length 10060;
Best Local Similarity	53.1%;	Pred. No. 1.1e-09;		
Matches 575;	Conservative	0;	Mismatches 480;	Indels 27;
				Gaps 4;

[illegible][illegible]

RESULT	4				
AB070947					
LOCUS	AB070947	24268 bp	DNA	linear	BC1 11-OCT-2001
DEFINITION	Streptomyces avermitilis polyketide-9 biosynthetic gene cluster.				
ACCESSION	AB070947				
VERSION	AB070947.1	GI:15824107			
KEYWORDS					
SOURCE	Streptomyces avermitilis DNA.				

REFERENCE  
1 (bases 1 to 24268)  
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces

**TITLE** Kluchnech, H., Shiba, T., Sakakura, Y. and Hattori, M.  
Genome sequence of an industrial microorganism *Streptomyces avermitilis*: Deducing the ability of producing secondary metabolites  
**JOURNAL** Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)

**TITLE** Direct Submission  
**JOURNAL** Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of Microbial Chemistry, School of Pharmaceutical Sciences, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan  
(E-mail: ikeda@mc.pharm.kitasato-u.ac.jp, Tel.:+81-3-5791-6242, Fax:+81-3-3444-6197)

FEATURES	location/Qualifiers
source	1..24268 /organism="Streptomyces avermitilis" /db_xref="taxon:33903" /note="Actinobacteridae"
CDS	1..1875

```

/product="putative membrane protein"
/protein_id="BAB6274.1"
/db_xref="GI:15824108"
/translation="MSAYHSPFIWAAVAVLVMTFVAVSRQNVHAPSGGDEYVANT
NIGPRAGILTVASALIVDVLVFAVSIAGSIENLGSAIPFVYVHKSCAAVAVLTLLT
NIRKRGESGKLPAITPYVFAVGFNIMAGARGVLDDTMAPTADYHINKHOGFL
GVALFELLRAFSSGCAALITGEALINEVSNPAKRSKNAASTLAMGLIATYMGCG
IALAAATVYRMAENPANDLHNGVAVGDPDYVNPVISOAVEAFKSGSLFVTLAAI
AVLEFLAANTVANGPFLSILAODYLPRLQHTGDRILAFNGSLVLLAGCATLLV
VCASDSTLIOLYIVGVFPELSLISQRYGRPHMRHRTLATEKDOKRHHVYSRAIINFG
PFTGVLVAVLVYTKFTHGAMVALLGMCIFYAAMSARKHYDVAEILAAPEGSDDS

```





overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Cosmid H63.

## FEATURES

## source

Location/Qualifiers

1..44274

/organism="Streptomyces coelicolor A3(2)"

/strain="A3(2)"

/db\_xref="taxon:100226"

/clone="cosmid H63"

/complement(1..957)

## gene

/gene="SCH63.01c"

/complement(1..957)

## CDS

/gene="SCH63.01c"

/complement(1..957)

/note="SCH63.01c, possible ABC transport system integral membrane protein (fragment), len: >319 aa; similar to TR:CAB77000 (EMBL:AL159178) Streptomyces coelicolor putative integral membrane protein SCH22A.31, 850 aa; fasta scores: opt: 510 z-score: 493.4 E(): 5.8e-20; 38.1% identity in 315 aa overlap. Contains possible hydrophobic membrane spanning regions"

/codon\_start=1

/transl\_table=11

/product="putative ABC transport system integral membrane protein (fragment)"

/protein\_id="CAC10292.1"

/db\_xref="GI:10432458"

/translation="MSLKNGSLRAAARVPRPAPFAGTFVALISALIVACGVILETS IRASVPAERYANAPYVADQASARVADVDDEPTATPLPTAVADGLAKKAGVP GAATVAPDPTFFVHGSDPSGALITGHGWSHATGTALSGAPAHAGEVVLGADARTA KAGVGDTVLETFAGSGAGFVSGLAEGAGDVGAGGATAMPADAEAPLADHPG QADALAVTAKGDADTDLAAAVEQALITGTAQTILGCDRGAEEDHGLAYGKETTFEAI GSEGIATLVAVFTAAAGTVALSVGQRTREFALLRAVATPRQI"

## misc\_feature

1..155

/note="nominal overlap with Streptomyces coelicolor cosmid SCH22A"

/complement(954..1748)

/gene="SCH63.02c"

/complement(954..1748)

## CDS

/gene="SCH63.02c"

/note="SCH63.02c, possible ABC transport system ATP binding protein, len: 164 aa; similar to TR:CAB81857 (EMBL:AL161691) Streptomyces coelicolor putative ABC-transporter ATP binding protein SCD40A.12c, 246 aa; fasta scores: opt: 1007 z-score: 1087.6 E(): 0; 67.4% identity in 236 aa overlap. Contains pfam match to entry PF00005 ABC\_tran, ABC transporter and match to Prosite entry PS00017 ATP/GTP-binding site motif A (P-loop) and PS00211 ABC transporters family signature"

/codon\_start=1

/transl\_table=11

/product="putative ABC transport system ATP binding protein"

/protein\_id="CAC10293.1"

/db\_xref="GI:10432459"

/translation="MDHREVPAAAPPGIAGVELRGVRRYRGAGTVAHLAGIDLALPR GFTAVMGPSGSGSTFLQCAAGIDRPSAGSVADLTGKNNENLELRSLRGFV FOAFNLPLPELVEONVILPMRLAGORHRAADVLAVGADKARRRRGELSGGQQQ RYAVNRLVLTADPVI FADDEPTGALDPTGTAAVTVGLLRQAVDGLGTYVWTHDPAAAA WADRYVFLADGAFADHLERSAQIAARMTALTGRATAAGVAA"

## misc\_feature

/complement(1062..1616)

/gene="SCH63.02c"

/note="Pfam match to entry PF00005 ABC\_tran, ABC transporter, score 208.40, E-value 1.1e-58"

## misc\_feature

/complement(1245..1289)

/gene="SCH63.02c"

/note="PS00211 ABC transporters family signature"

## misc\_feature

/complement(1572..1595)

/gene="SCH63.02c"

/note="PS00017 ATP/GTP-binding site motif A (P-loop)"

## gene

/complement(1957..2631)

/gene="SCH63.03c"

## CDS

/complement(1957..2631)

/gene="SCH63.03c"

/note="SCH63.03c, possible two component system response regulator, len: 224 aa; similar to TR:Q9X807 (EMBL:AL049754) Streptomyces coelicolor putative two-component system regulator SCH10.18c, 228 aa; fasta scores: opt: 827 z-score: 946.4 E(): 0; 62.1% identity in 219 aa overlap. Contains pfam matches to entries PF00072 Response\_reg, Response regulator receiver domain and PF00196 Gere, Bacterial regulatory proteins, luxR family and match to Prosite entry PS00622 Bacterial regulatory proteins, luxR family signature"

/codon\_start=1

/transl\_table=11

/product="putative two component system response regulator"

/protein\_id="CAC10294.1"

/db\_xref="GI:10432460"

/translation="MTTRVLIIVDDQAVRAGFALLAAGSDIDVGEAPDAGQVELS RRTHPDVLDVRRPMDGELARRLSPPGVYHRRVRLMTFTFDVDDVYALRAG ASGLINDADPADLILAAVRVAGSDALLASVTRRLIADPAQGRPASRGAPRLRLKGL TERTEVLTIVARQSNTEIARTLVLAEGTVKTHVSVLRKLDLRDRAQAVPAYSGL LVARGF"

/complement(1966..2163)

/gene="SCH63.03c"

/note="Pfam match to entry PF00196 Gere, Bacterial regulatory proteins, luxR family, score 83.30, E-value 4.9e-21"

## misc\_feature

/complement(2029..2112)

/gene="SCH63.03c"

/note="PS00622 Bacterial regulatory proteins, luxR family signature"

/complement(2272..2625)

/gene="SCH63.03c"

/note="Pfam match to entry PF00072 response\_reg, Response regulator receiver domain, score 124.80, E-value 1.6e-33"

## gene

/complement(2628..3974)

/gene="SCH63.04c"

/complement(2628..3974)

/gene="SCH63.04c"

/note="SCH63.04c, possible two component system sensor kinase, len: 448 aa; similar to TR:Q9X806 (EMBL:AL049754) Streptomyces coelicolor putative two-component system sensor kinase SCH10.17c, 472 aa; fasta scores: opt: 595 z-score: 616.4 E(): 8.2e-27; 36.4% identity in 453 aa overlap. Contains possible hydrophobic membrane spanning regions"

/codon\_start=1

/transl\_table=11

/product="putative two component system sensor kinase"

/protein\_id="CAC10295.1"

/db\_xref="GI:10432461"

/translation="MRSRRRLPRRRGARRAPRLVMPIVEARPAIILAGSRMY RLPMVAFLVCLALPPTIVLTADYGLNGSLASALAVQAAPLLAVYRPLQAMTY VEVADVAGALALTLVDFQERILMPFPMETLVGVGLCLAGLRLERTLLVLAIRAG ANVLGDAAPRGITGAKDLTLILGVALILGVALRLERYEVRRLARDETLLSEMERGR TILDERARILRELDVVAAHNSVTYVADLGAALRLAPDVDEEFTSIATARESLG EMRLLELRLRENEHGLVLPQPGALQIGLVATARVAVPVGDCDAEAPVEVAGSA YRIYQELAVNVRAPCAATHVSVARPPACGCGCGCGTODRRCGGERLTIVYNEP PPEPPAPLEGVGTGHLVGRREVRVLVGSTLVDGLPDDGFRVAQPLTEEDIT"

/complement(2639..2642)

/gene="SCH63.04c"

/complement(3983..3987)

/complement(3987..5039)

## RBS

## gene

## Query Match

## Best Local Similarity

## Matches

## Conservative

## Mismatches

## Indels

## Gaps

## 4;

## 27;

## Gaps

## 145

D	b	16134	CCGTGGGCGCCGAGACCCGGGTTCTGTCGCCGGGACGCGGGGAGACCGCGCTGGTGGTGA	16193
Q	y	146	tgcgcgcygttgcctcgagcagaacgcttcgccggaaaccgcaatccgcygtttcttgaaaccg	205
D	b	16194	CACGCGAACCGCACGGGCGCCGCCTCACCAGCAACCCGCGCTGGCGAACAACATCCGGCACT	16253
Q	y	206	aacctcgtcgagcccgaggcctcgcacgcgaagtatcgttcggaactaatctgttccaagagc	265
D	b	16254	CCGCTTCCCTGGGACACGCAACGGCGGGGCACACCATCGGGCCACAACATGCTCCAGAGCGAAC	16313
Q	y	266	agccgcagcaaacgcgcgcctctgcgcgggtgttgtcgcgcgcgcgttccacgcgcgcgcgcgc	325
D	b	16314	CCCGGAGACACCCGCTGCGCGCGCGGCTGTGCGCGGACACTTTCACGCCGGGACGACCG	16373
Q	y	326	gccgcgctgaacaacgltgatcgcgcgcgaacccgtgagcaacctgtctgcgcgcgcgcctgcgc	385
D	b	16374	CGCGCCCTGGCGCCCGCGCGCTGCGAGCCATCCGACACGACTTGCTGGACCGCTGCGCGCG	16433
Q	y	386	ggggcgagatgtagcgttgttcgagacgcgcgttaacctgtgcgttcgcgttcgcgttcgc	445
D	b	16434	CGGGACAGGCGCCACCTGGTGGCCCGGTAACGCGCTGACCGCTGCGGCTCACAGGATCTACG	16493
Q	y	446	gscctctgcgcctctgc	505
D	b	16494	AACCTGCTGGGGCTCCCGGAGAACGACGCCGAGAACTTTCACACCTGGTGTGAAAGAGCTGG	16553
Q	y	506	gaacggaacctggaacgc	565
D	b	16554	TCAATCCCCCTATTGCGCCCGCGAGCGCGCGGCTCG-----GGCGGACCG	16595
Q	y	566	cgcgagcttcgc	625
D	b	16596	CCCTAACCGGCTTACCTCACCGGCTGTACGAGCCGCAACGCGCGGCACCCCGGACGGCAAG-	16654
Q	y	626	aacctctgcgcgttgaatgtctgcagacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	685
D	b	16655	--CTGCTGGGCGATCTGGGTGCGCGCGCGCGCGAGAGGGGGAATCACCCCGGCAACTGC	16712
Q	y	686	tcaagaacggttgcacgttatcatlcttacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	745
D	b	16713	TGGGATGAGCCCTCTCTGATCTCGTGGTGGCGGGCACGAGACGCACTCAACTGATCTCG	16772
Q	y	746	agcgcgtctgcgcctctgc	805
D	b	16773	CCACCGTGCACGGCTGCTCAACCCACCCCGCGCACTGCGCGCTGCGCGCGCGAGCCGG	16832
Q	y	806	aacctctgc	865
D	b	16833	AGCTACGAGGCGCGCGTGGAGGTGCTGGCTGATCACTCGCGCGCTCCACGCGCTCG	16892
Q	y	866	cccgc	922
D	b	16893	CGTTCCGTTTCCGCGCGCGCGAGCGCTTGGAGTGGCGGCGCACGCGATTCGCGCGCGGACNC	16952
Q	y	923	tcgttgtctgc	982
D	b	16953	CGGTCTGCTCTCCCTCCCGCGCGCGCTCCCGAGCCGCGCACACTTCCCACACCCGCGACC	17012
Q	y	983	attcgacatcacgagcgagatccggttccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1042
D	b	17013	GCTTGGAAATGCGCGCGCGCGCGCGCGCGCGCGCACTTGCTTGGGACACGCGCGCGCACT	17072
Q	y	1043	gocctcggttctacacctgc	1099
D	b	17073	GCGTGGGCGCGCGTGGCG	17132
Q	y	1100	gactgc	1159
D	b	17133	GCGACATCCGCGCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	17192
Q	y	1160	tgcttcgc	1194
D	b	17193	TGCTGCGGGGCTGTGGAACTGCGCGCGCGCGCTC 17227	

	RESULT	6
AL646080	LOCUS	
DEFINITION	Al646080 190050 bp DNA linear BCT 07-DEC-2001	
ACCESSION	Ralstonia solanacearum GM11000 megaplasmid, complete sequence;	
VERSION	segment 5/11.	
KEYWORDS	Al646080 Al646053	
SOURCE	AT646080.1 GI:17431113	
ORGANISM	Ralstonia solanacearum.	
REFERENCE	Ralstonia solanacearum; beta subdivision; Ralstonia group;	
AUTHORS	Bacteria: Proteobacteria; Ralstonia.	
TITLE	1 (bases 1 to 190050)	
JOURNAL	Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S., Ariat,M., Billault,A., Brottier,P., Camus,J.C., Catolico,L., Chandler,M., Choisme,N., Claudel-Renard,C., Cunnac,S., Demange,N., Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schlex,T., Sigulier,P., Thedault,P., Whalen,M., Wincker,P., Levy,M., Weissenbach,J. and Boucher,C.A. Genome sequence of the plant pathogen Ralstonia solanacearum Unpublished 2 (bases 1 to 190050) Boucher,C.A. Direct Submission Submitted (03-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS, BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS 118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA URVY, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biométrie et Intelligence Artificielle INRA, BP27, F31326 Castanet-Tolosan Cedex, Laboratoire de Génétique Cellulaire INRA, BP27, F31326 Castanet-Tolosan Cedex Christian.Bouchere@toulouse.inra.fr <a href="http://sequence.toulouse.inra.fr/R.solanacearum.html">http://sequence.toulouse.inra.fr/R.solanacearum.html</a> .	
COMMENT	location/Qualifiers	
FEATURES	1..190050	
SOURCE	/organism="Ralstonia solanacearum" /plasmid="megaplasmid" /strain="GM11000" /db_xref="taxon:305" 51..362 /gene="RSp0643" /note="RS05579" 51..362 /gene="RSp0643" /note="RS05579"	
gene	/gene="RSp0643" /function="miscellaneous; unknown" /note="Product confidence : hypothetical Gene name confidence : hypothetical predicted by Codon_usage predicted by FramedP /codon_start=1 /transl_table=1 /evidence-not experimental /product="HYPOTHETICAL PROTEIN" /protein_id="CAD17794.1" /db_xref="GI:17431114" /translation="MKTPSPNPAAGPASDPKAGDGVAPEQLHEYNHTITPR RIGGGIMRAFAVSLDSRALVAVTGVGAFFPIGARDAVDANSIFIDRLKACP A"	
CDS	397..555 /gene="RSp0644" /note="RS05578" 397..555 /gene="RSp0644" /function="miscellaneous; unknown" /note="Product confidence : putative Gene name confidence : hypothetical predicted by Codon_usage predicted by FramedP"	
CDS	gene	

gene

/codon\_start=1  
/evidence=not\_experimental  
/product="PUTATIVE SIGNAL PEPTIDE PROTEIN"  
/protein\_id="CAD17795.1"  
/db\_xref="gi:17431115"  
/translation="MERQQLFAITVAIVAAVSAVSLANDLNDERDPPGQAALVHGH  
ATPSAAGR"

gene

complement(575. .1135)  
/gene="RSP0645"  
/note="RS05577"  
complement(575. .1135)  
/gene="RSP0645"  
/function="miscellaneous; not classified regulator"  
/note="product confidence : probable  
gene name confidence : hypothetical  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/transl\_table=1  
/evidence=not\_experimental  
/product="PROBABLE TRANSCRIPTION REGULATOR PROTEIN"  
/protein\_id="CAD17796.1"  
/db\_xref="gi:17431116"  
/translation="MKRISTPDPFVDMKKVAIKRSPAAAPGGAALSAIDRDRATLL  
KMLQDASISNVALADKYSLSPAICRIVPEIKELGCRGVALLERKALDGLTIVT  
GLVDNRSPESFADPERKAVQKPPGCLGEHVYTGFDIFMLVTRDNDSYNRLAEQLL  
VLPGRQRIREFITLAKQVLSLTQPLA"

gene

1244. .2260  
/gene="RSP0646"  
/note="RS05576"  
1244. .2260  
/gene="RSP0646"  
/EC\_number="4.1.99.4"  
/function="miscellaneous; hypothetical/global homology"  
/note="Product confidence : probable  
gene name confidence : hypothetical  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/transl\_table=1  
/evidence=not\_experimental  
/product="PROBABLE 1-AMINOCYCLOPROPANE-1-CARBOXYLATE  
DEAMINASE (ACC DEAMINASE) PROTEIN"  
/protein\_id="CAD17797.1"  
/db\_xref="gi:17431117"  
/translation="WNLNKRPFRPPLTGPPPTIQPLKRLSHLGGKVELYAKREDCSNG  
LAFGNGKTRKLELYLPEVLAGGYDTLVISIGISQNTOROVAANAHLGLICVCYVQEWX  
VNAADVADVRNGNELSRILGADVRLDAAGFDIGIRSWEQAMEDVVRAGGKPPIPA  
GCSHEPDLQGVFGFAEVRQGEAEFGFRFDYIYVCSVTSGTAGMNVGPAAGRAR  
VIGIDASKAPQTEQRIILRIARPAKLVLEGRODITVEDVLDFTYGGPEGLNEGTL  
FAIRICAAQEQELTDPTYEGKSMHGMIDVRGSGFPFGSNTVLAHLGGVAPALNAYSTL  
FRNG"

gene

complement(2336. .3508)  
/gene="RSP0647"  
/note="RS05575"  
complement(2336. .3508)  
/gene="RSP0647"  
/EC\_number="4.2.1.17"  
/function="miscellaneous; hypothetical/global homology"  
/note="Product confidence : putative  
gene name confidence : hypothetical  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/transl\_table=1  
/evidence=not\_experimental  
/product="PUTATIVE ENOYL-COENZYME A HYDRATASE PROTEIN"  
/protein\_id="CAD17798.1"  
/db\_xref="gi:17431118"

CDS

CDS

CDS

			/translation="MSALPKPAEPAAAAAEPEVLFFHVNRVALITLTNPRESALTNALSHG VQSLALLAQCRDDPRRIAAVVLYRGGEFGFCGGGVRLATLRYLRGOAHAGDGMLOF YDEIETALHREFPKEPVVALLDGIAMGDGLGCAGARLVREYSRIRAMPETRIGFL DYGTAFILCAMPEEMALYTGLTTLGSADMGICGAADVPCVAEMLASFEDRLISLE AGDPDIOMLARVRVEPPCNIIVPHAIARTTPMYRVRRDRSTIDRIYAQLRESIERC PAREROWILOATLMATHSPMTMHLVTSEKLLRGOUTLABCFRNELCIVARATEEGC FEGBRAHDVDHDRHPGNAPATIACVGVRPERVRHFLLSPWRCAGHAHPLAHLQAOPAA" complement(3505). .4305)
gene			/gene="Rs0648" /note="RS05574" complement(3505). .4305) /gene="Rs0648"
CDS			/EC_number="4.2.1.17" /function="Small molecule metabolism; degradation; carbon compounds" /note="Product confidence : probable Gene name confidence : hypothetical predicted by Codon usage predicted by Homology predicted by Framed" /codon_start=1 /transl_table=11 evidence=not_experimental /product="PROBABLE ENOYL-COA HYDRATASE PROTEIN" /protein_id="CAD1799.1" /db_xref="GI:17431119"
gene			/translation="MALDWLHDGTVALTTLSRPANAFEPEGRLQOLAVETLEDABE RYRALVIITGDGPKEFSRADNTEFASGRDVARERAAREGAEFALOVARVIAAILIN GYAWMGGLECALACDIRIAERNHQLAVERAYGLCCCGQTLPWLIEBMARKMITI TGERIDAOTALARIGLIVEVPDTGARREALMARVTVTSRAVFESALTHQRHGCV PTATAVEREREFDVLFHPDOREGVANFLEKRPMHTARPBEALP" complement(4302). .5198) /gene="mmsB" /note="RSp0649; RS05573" complement(4302). .5198) /gene="mmsB" /EC_number="1.1.1.31" /function="Small molecule metabolism; amino acid biosynthesis; isoleucine/valine" /note="Product confidence : probable Gene name confidence : putative predicted by Codon_usage predicted by Homology predicted by Framed" /codon_start=1 /transl_table=11 evidence=not_experimental /product="PROBABLE 3-HYDROXYISOBTYRATE DEHYDROGENASE OXIDOREDUCTASE PROTEIN" /protein_id="CAD17800.1" /db_xref="GI:17431120" /translation="MKTAIFGISLGMMGMAPMNLIKAGATSVPDLDAVOSVIDAACR SAASPRAAKTAGCACCAITMPLPAAHHVSULTDADGITLAGIKVPIIIDSSIIIPDSKA ALAIALTRHNPFYPDALPYSGCTGGAAGTILEFMGGANAPEOVPLVASAGKNLVC GEAGAOGGARICTNNLVLIITMAGVAEMSLEGALIDPKVLGGIINSTGCMWSDDTY NPFPVELITEASPSRGRTGFCDTMLKDILGADGAANSVROPYYLGAALOULOASSSS QGACKDFESAISRILRYPGAAGR" complement(5220). .6767)
cds			
Query Match	17.8%	Score 215.6;	DB 1; Length 190050; Best Local Similarity 52.2%; Pred.No.7.le-10; Matches 529; Conservative 0; Mismatches 479; Indels 6; Gaps 2;
Dbs			
Db	77566	GCATCGCGTCGGGANACGCCGACGAGAGCCCCACAATGCCGTGTTCACAGGGGTTCACC	77625
Qy	245	aacatcaactgttcaggagaccgcggaaccccgctgccggagtgttgcggcccgc	304
Db	77626	GGATGTTCTTGATTGTCATATCCGCCACATGACACCGCGCTGCGCGCTGTGATGACGAGG	77685
Qy	305	cgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgatcgcgcgcgcgcgcgcgcgcgcgc	364

Dh	77686	TTGTTACACGGCGGCAGATTCGATTCGATGTCGGAGAGTCCACACGGCCACCGGCATCACT	77745
Qy	365	TCCTTCGCGCCCGCCCTTCGCGCCCGGCGCGATCTGAGAGTGTGTCGACGAGCTTCGATCCCGC	424
Dh	77746	TGATGACGACGACTTCGAGCAGAAAGCCGTCGGCCGATCTCTGCGCGGAATTCCGCTTCCCCC	77805
Qy	425	TCGCGCTCGCGCGGTTCTCTGCGCTGCTCGCGCTTCGCCGCGCGATCGGCGGCGGCTGCG	484
Dh	77806	TTCCCGTCCGGATCACTCTCCGACAGATGAGTGACCTTCGATATCGACCATCTCCATGCGCGTG	77865
Qy	485	GGCGCTTGTCTCGCGCACTTGGAGACCGCTTCGACCGGCGCGC---CAGCGCCGAGGACA	541
Dh	77866	GCGTGGGGGTGAGCAGAGCTCGCAGAGGTTCGATTCGCGGCGCGATGAGCGCGCATGCGC	77925
Qy	542	TCGCGCGCGCCACGCGGCGATCGCGAGTTCGCGCACTGTTGAGCGGCGCTCGCA	601
Dh	77926	TGTGTGAGACGACGAGCCCGCTTCACGAAAGACTTCGCCCACTTCCAGAAAGTGATCGAGG	77985
Qy	602	GGCGCGCGCTGAGAGGCGGCGAGGACCGCTTCGCGTTGATGCTCGACGCCCGACGCGCG	661
Dh	77986	CCCGCCGCGCGACGCCCGCACCCGACCGACTGATTTCATGCTGATGCGCGCCCGAGGAAAAAG	78045
Qy	662	GCCTGA---TGAGTCTGCAAGAGATCGTCAAGACGAGTGTGTCAGTTCATCTCAACCGGCG	718
Dh	78046	GCGAGACCTGTGACGCGACATGAAATCGTGTCCAAAGTGATCCTGCTGTTTCATCGCGCGCG	78105
Qy	719	ACGAGACGGTGTGCGCGCGCGAGTGTGTGCAACGCGGTGTGAGCTGTGTGCGACCCGGAGC	778
Dh	78106	ACGAGACCAACGTCACCAATGATTCGGCCAAATGCGCTATGCGTGTGATCGCAATTCCGCAAGC	78165
Qy	779	AGCTCGACCTGTCTCGCGCGCGCGCGAGACCTGTGCGACAGCGCTGAGAGATGTCTTC	838
Dh	78166	AGCTCGACCTGTCTCAAGCGCGAGCGCGTGGCGGATGCCAACCGGTGTCTGGAATCCCTGCG	78225
Qy	839	GGTACGACCGTGTGTGCAATTCAGTTCGAACACCGCGAGCTCGAAGTCAAGTGTGAGCTGCG	898
Dh	78226	GCTACGACGGGTGCGTGTGACAGTTCAGATTCGCGCGCGCTGTGAGGACGTCGAAATGTCGAAG	78285
Qy	899	GTCTCGCGCTGTGCGCGCGAGGATCTGTGTGTGTCTGTCTGTGCGCGCGCGCGAACC	958
Dh	78286	GCGAGGTGCTGCGCGCGCGCACGACGAGGTCTCTATGCTGTGCGCGCGCGCAATCCGATTC	78345
Qy	959	CGCGCGGTCTACGACCGCGCGCGACGATTCAGACATGAGCGGATCGATCCGCTCGATGT	1018
Dh	78346	CGGCGCAGTTCACGATCCGAGCAGCTCGATATCGCGCGGACGACGAGGCGGCTGTGCAGA	78405
Qy	1019	CCCTTCGCGCGGAGATGTGCTACTGTCTCTCGGTCTCTACTGCGCGGTACGCAAGCTGCG	1078
Dh	78406	CGTTGCGCGCGCGGATCATCTTCTCTCTCGGTATCGGCTCCGCTGATCGAGCTCGAAT	78465
Qy	1079	CCGCGGTGTGCGCGCGCTGCGCGAGCTGTGCGCGGCGTGTGCGTGTGCGGTGTGCGAGCGCC	1138
Dh	78466	CGCGCGTGGGCGGCTGTTCGAGCGCGCTGCCAATCTGGCGCTCACCAATCTGACACGAC	78525
Qy	1139	TGGCATATGAGCGCGCAACATGTCTCGGCGGCGTGTGCGCAAGCTGTGCGATCGCGT	1192
Dh	78526	TGAGCTGGAACACGCGGGGCACTGTGCGGGCGGTCAACGCGCTGATGCGGCGCT	78579

[illegible]

ACCESSION	complete cds; and unknown genes.
AF127374	AF127374.1 GI:4731328
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	Streptomyces lavendulae. Streptomyces lavendulae Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE	1 (bases 1 to 63734) Mao, Y., Varoglu, M. and Sherman, D.H. Molecular characterization and analysis of the biosynthetic gene cluster for the antitumor antibiotic mitomycin C from Streptomyces lavendulae NRRL 2564 J. Biol. Chem. Biol. 6 (4), 251-263 (1999)
JOURNAL	99201491
MEDLINE	2 (bases 1 to 63734) Mao, Y.O., Varoglu, M. and Sherman, D.H.
REFERENCE	Direct Submission Submitted (10-FEB-1999) Microbiology, Uni. of Minnesota, Box196, 420 Delaware St. SE., Minneapolis, MN 55455, USA
AUTHORS	3 (bases 1 to 63734) Mao, Y.O., Varoglu, M. and Sherman, D.H.
TITLE	Direct Submission
JOURNAL	Submitted (27-MAY-1999) Microbiology, Uni. of Minnesota, Box196, 420 Delaware St. SE., Minneapolis, MN 55455, USA
REMARK	Amino acid sequence updated by submitter
FEATURES	Location/Qualifiers
source	1..63734 /organism="Streptomyces lavendulae" /strain="NRRL 2564" /db_xref="taxon:1914" 123..1367 /function="dehydrogenase" /note="ORF6" /codon_start=1 /product="unknown" /protein_id="AAD28446.1" /db_xref="GI:4731329" /translation="MRTNIFELGDIPIPTVPEQMVASVIRREYRGPHAFRSEVVD VPKPGQALVLMVAGINYNVNASIGQPVDSIAHOKGSHDFIHGSESGGVVW AVGSGVGVNAGDEVITLSCGQMETADIRLGDAPMTISGSQVWVGEGNGSFPAPNL VDDVQHKRPGLTWEEBACFLILGATATATRCGQWQHDARPPQVILMGAGGLGS AIQITTRAGGIPVAVVSDEERARYCRELGAGTINRLDFDHGRLPDIGHEANGKMT EGVAFGRREWEVLGERSPRIVLEHSGQAVIPSMILCDAGVAVICGGTGTGNADIG DRLFMKROKRLGSHFANLRQCRDVIHMVANGDLDPCLSWTGGFDDIGKAHQLMHDN OHPGNOAVILVNAFRGTGTTFA" 1428..1958 /note="ORF5" /codon_start=1 /transl_table=1 /product="Lina homolog" /protein_id="AAD28447.1" /db_xref="GI:4731330" /translation="MDKIDILSERETIRAVLQRCRCGLDRLDELIVKSAHYHEDAHDDR GVINGNAHDFVKQIPLRLDVAITGLTLTHSTLEIDGDAVGSTCTAHTRSDEGI KRVQVAGRYVDFRERDGVWVIARRVLNDFSLAOEVPDIPAAQAGFNPSHRDLDTID ASYOVLPTRGPDAPTL" 2006..3229 /codon_start=1 /transl_table=1 /product="cytochrome P450 hydroxylase ORF4" /protein_id="AAD28448.1" /db_xref="GI:4731331" /translation="MSDTEOHAPLTPRQRCQPSPPLELELRTDPTSRMFADSDS GMLTRHADVRAALADREVSSHPKAKQRPNNLAPEKRAEHTLPGFLFMPPRPTTKR RRLLTKEFMTRAIKRLKEPRLEQIVTETLDADAGGVVDVQSFALPLPLVLTLELKG IRYEREFEMDMVLRQLADATPELAGLARGMEFMKLAARANGDGLSHLHNA DPDADPALTDEIAGIYMLIAGHETSANMLGVTYTLLENADQMLLRDIDSLIRB AVELLRHOTIVQOGLRGVTRDMELINGHOVKTGESLASSPANRDPVAPDPDRDQ ITREHNPLATGCHSILCLGMLARVEMRQAMRVLVTRFELRMAANPEDIRMDDQ VYGVYLVPTVDEAK" 3226..4458



gene  
CDS  
gene  
CDS  
CDS  
CDS

```
/codon_start=1
/transl_table=11
/product="cytochrome P450 hydroxylase ORF3"
/protein_id="AAD28449.1"
/db_xref="gi:4731332"
/translation="MTGPEAAVRGCGPFGAGAPAPVPHFAPDRLEPDPEWPELRERPL
ORVTLPGGEMAIARVODVRAVADPREFSQOLAVAGAPREFLEPHQPPDVAIVSEGP
DHARLRILYGVYFTPRVEDMRPLIQRADGLAMEBMPADLVEDFSLPFVSVTI
CELLVAPPEDEKRFRCVWSDALLTTTAHTPAVVRVMMQMDHYLGGLVAQQRVETAD
IGSLVADAEEDKRLTEGELVRLAEILITAGETSASQIPNLTLYLFPHQLTEIRND
HDILDAVEAEILREYPLIGTVGDFPRTAEDEVLELGSVLRACEIVVPSMGANRPLEF
TDPDLDLARRRPNPLIGAGAPRHNLQALRAVELQITLTTLFRRYRRLRLAVEESL
SMKEIAMWGKHTMVTW"
4552..5658
/function="esterase"
/note="ORF2"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD28450.1"
/db_xref="gi:4731333"
/translation="MSTIDEWHSKTEAGMDPALRLRTIDVVRAGGAOLCVNRGRT
VLDLSCFGSSDLSFLVYAATRPVAAALVAHALERGLIDRPVAEEWPFARIGKGD
VTVRHVLDHRAGVAVGRIGVATKTRPAALGAMERSVADLEQSRKMGGEVAAVHNSFCF
ILGELVQVQAVTSPEDFEVTSLSLPALGLINDHMLGPSAMVRRVRAAHADSEMPNO
MSNRGYGQAVTIPSGSLGSTAQAOMARFQOMMEGGSIDGIIVLRPEVTEARKRSNG
GIDSLKRPVMSHGFMLGGCPGPPRGISLVNGSTSDPSAFGHAGTTSVYVADPTRE
LYALSLNIOEPFGAGTIRLKEVSDLAIGCEAG"
5992..6849
/function="transcriptional activator"
/note="ORF1"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD28451.1"
/db_xref="gi:4731334"
/translation="MLNPKMERANRHPSPQVGLCPLEVRSAGAGTGAAGSIVIR
RTLLAALRLAIGQAMSTERILKEVWADNPATDTRKAAVAVALRLRYLGNEGSWILT
APSTVSDITPELDVAVEALILVREGRAALAGAPVAAVRLTALOOMREPPADLTA
AISYSCSTIPEENLRSEAVOAHIDRALEIHHOELVGEIRSLNAPLNEPHILOAM
IALYSGGQAEMLAAVMOALRAENLGDIPGRQLOQELHLRLIRADNGLITGSPAP
AEPLVROS"
complement(6854..7666)
/gene="mlt"
complement(6854..7666)
/gene="mlt"
/function="aminoguanate dehydrogenase"
/codon_start=1
/transl_table=11
/product="Mlt"
/protein_id="AAD28452.1"
/db_xref="gi:4731335"
/translation="MEFLPAGPAGVSGATRLVAVLDVPAOVKAPGLNPLISEGDLA
VVVPHAVARLDALEVYEGELKRGINDGLILVYVPRKALGGLADLGRNALIGANMA
REDEGRKXANPEFGSLGCVQSLQAAGHTVRRRNVLYGAGAGSALITALLMADANV
SVNHTDRAQDLRLRGSRRPDGIRALGEPDLEAHPFANVATPLGNRSDDPPEFDP
RVRPVAVVADVMMRPHETALLSAAATGRRVNHGIMLEDDQVPCYRAFFGMP"
7895..8842
/gene="mlt"
/gene="mlt"
7895..8842
/gene="mlt"
/function="kinase"
/codon_start=1
/transl_table=11
/product="Mlt"
/protein_id="AAD28453.1"
/db_xref="gi:4731336"
/translation="MTGPDGAGGGGVYTFRRPAGVDVYADLDLAAVRLGTHREPP
LAGVAVAPATCDAGYVYRTWPRSPNAGLNLITAEQQLRPLGTVPVACADGGDLAAE
SRAACRKLIVYGTGIGGGIVHEGRAMPBPGHSCVEGHVHVVRDPSGPKDCGRAC
VQAAVSGALTLRRAAERGRGTGDELASGARLRAWPAEAAVDSAAALTAVYIGICE
LAHPLIVYGGGFAAGYGVYAAVAAHYERLIRGTDPVVRVAVYLGGRSILHALL
AREALGRNRPESRVSVDVSPVSGYVDFDRAVGRSD"
```

[illegible]

[illegible][illegible]





FVPSMTLPFLDLDDPTARAASSLRLVFCSGEALPLQVAFRRLLFGDAVNLVLY  
GPTFATVDSDEHCAADNPTRVPIGRITDNLRLLVLDRLRPOPLGAVGLYIGAV  
ARGYLNRPNLNERPLNDPEVAVGGRLYRTGDLRLADGNLLETIGRADPOVKIRGRV  
EPDEVDRRLALAGVDAVAVARDSAGVTHLYVYAAAEIDRGORLNGSLSTLPDF  
MLDAPFVRIDSLPLSANGKIKDRRCI.PAPREVAVAVAPRTAEALRNADVYGVAV  
GVHDPFYALGDSITMLRIIRAAGRCIGFELADLMNPVTAGLAEPLRLPLAKERYO  
PEFLVSVDRPRLIEGLEDAFPTRSLSTLGLHSORPDSVSYHDVTRPDLMAADIA  
FRHALDRVAAVAPALRSSSDLSGASEPLQVHTHAKSEPLILDRGPEGVYLDHIE  
RORRFRYSILOQPGLEFLAAVREDGDLVVSFHHAILDQMSVANLIVAAVAREP  
LPGPAPALACHVREELALASPAAGVYTGLEGARMTRLDGFAGNEPQAPASHR  
EALPDGLERLKATAORGLPLKSLLLAHCITLFLFRSDSVYTGASISRGRELPA  
DRMWGLFNTVYVSEIAGCSWIEVDALPFOERDGHARRYPPLSAIOQIVGSELSSA  
FNVNTHVLEPLMOLRDRPVEETPMLVNVITAPSPDCMWTIRLIDSGEELSSOAL  
ICATPELLRLADHPDEADPAFLAPRDAASOPEPLVDSVLSFEQVREBALGSAL  
AFEEQWNTYRDLIDHARCVAITLVAARKSDAGVALNRSPEITATIMWILGAGVLC  
VPLDVSYPQORLALILETAQPRVAVPHEHNAVAARVYLVEELVADILEPTEFAQ  
LDELAMLLFTSGSTGRPKGVLSHRMANYOMOLRYASGVPGILTQJOFAPLSFDMAF  
QEIFSTLCCGGELOLISNERMDPSALLVLERQOVOLVLEVALORLAESANLGA  
RPGALRVVYSSGSEQLITEEDVRAFCAPAGLLLENQYGETHQVYHSLSGDPHAYP  
DLPPIGRPLDGVYVDALALBPVGVYTGELYFGDCLARGYHRAPKLTAERFVHP  
MRGALYRTGDLGRILNGEIVYVGRADPOVKYGRPRIAPVETELAIMORAROSGL  
RGAAYVARRQGNDAFLAFLILGEPEANDLAELOALRSELPHEMYPANRANVDFAL  
TPSGKRDALRALPLUEHGTNLETLPADODYERLGLAGLGLBLDRPVGTRDSFQDQ  
GTSLSAMRPMILLERKYGVDPMAALLETPTVEGLAEERLSAVRAFDLPVIRAGG  
SRPPLFLVHPGLGSHVLCYLPVLRALPPDOVYALQAGTGQGSTPLAVLBDIAASYLA  
AIRVOPEGPYLYGWSFGFVAYEMAROLRALDPOAVOQLYLDSTVYDRNAGSAS  
DEALLFEFWELWMPERSDKEVEPLPEGASLEOKLIDHIVERALPAGVLPAGTRAVO  
RLYELFRASMOALIGYRPEVSDPODWTILRADGPLALAKMHDAAGHYDPPANGQH  
WTSGRLDVITDYGDDHVLAKPEYVETVAEIAELALLEPSTSETRP"  
8344..9510  
/gene="PA3328"  
8344..9510  
/gene="PA3328"  
/codon\_start=1  
/transl\_table=11  
/product="probable FAD-dependent monoxygenase"  
/protein\_id="AAG06716.1"  
/db\_xref="GI:9949459"  
/translation="MSRHPLKIVVINGAGISGLAAACIKAGFEVELYERARELRAVG  
SALSLPMNALITALERYVRLDTRAQAFDSLFLTRGRPIRAIDEGFLAROLGOSL  
AHRASLQOALLEOARDCEILGVSATYLRHADGEGVYVCSGDRVHADVIGDG  
FNSAIRATWTGPERPTDMHYVIRATPFRHRYTPGVYAHYWGRCRFLADIGGN  
VYMWGTRNMPABOAKDMRGKAGIORLAGNADVOAVTEATPEADISLSPADREPL  
ERMGDGPVTLIDDAHPMLTSGOGAALAIIDRAVIAHCLATIDDDPALAIRRER  
DRAAMVETSRALSRIEQLNHLRTVADLYFRAPERTARQNELALTPGYE"  
9507..10835  
/gene="PA3329"  
9507..10835  
/gene="PA3329"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAG06717.1"  
/db\_xref="GI:9949460"  
/translation="MAAIRPLSLAVERWYLSDOFSALNVISRVYHGLSTIDLR  
GUDALQARHPLRLARLEHDAGLDPKRWPCERPILPREVRGGEQWMLREINELPER  
IDPDSGPLRTVAIATADAGADLVVYHIIADSTVYTLTAQMLTLAAPAOPTA  
SALPAPEDLIRPRPTGDEGAARLEAOTQADBALGRHRRGPESNPLPEARBTLL  
HRELDAQEOLORARAREHGTTHGALTAALAIAGHDHORPSHIIAGSDITREL  
EPVVRPDEVGTVAAYVYVLDIARPFVVARALINDDGERRROGHHNLTIVLASAP  
RCMADARPMWAEAEAGPINCSSNIGVYRPERIGALISDNOFLTGISVNYTPAA  
INSSHGRLFMNFFYIDEAVPGRBARLAEDELGLTLLSIHAPORSALSEO"  
10837..11751  
/gene="PA3330"  
10837..11751  
/gene="PA3330"  
/codon\_start=1  
/transl\_table=11  
/product="probable short chain dehydrogenase"  
/protein\_id="AAG06718.1"  
/db\_xref="GI:9949461"  
/translation="MLPTSKPLSPQGRVILITGASSGIGRETALHLAEQFOVIACVR  
ROEDGERLANACPSGRISTLLIDVTDEESIGRAAAQVAERVGTGIMGLVNNAGICIS

APLECVSDLLRROLEVNLIGOLAVTRALPLRLRGGAARLVNVTSGLSVAIPYLA  
YSMAQAKKGVSDALRLRELPAGIOVSVSGALWMPIMWKIASSEERLADAPAVA  
DLYRDIYLFELIANDGAGNSATKPDVAAVAAHVAALTAAPRFRYRVGADVRGCTLLA  
RLPDSIDGMGRPIVTTAAPAKKEQRA"  
11744..13000  
/gene="PA3331"  
11744..13000  
/gene="PA3331"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome P450"  
/protein\_id="AAG06719.1"  
/db\_xref="GI:9949462"  
/translation="MPDKRLRLEGEELISPLHATVDYGVNCGAPRAHRAEHVWVT  
RDKARKVNLHNGVRDARQAEELTAKTGSFPRAGTEGSLSHMLNLDPPDTRRLSL  
VGSAFTPKOVELOPHIRITETALDAAAGEQADLMADPAIPLTIAVIFELGLPEA  
ERBHAQADQAEVROAELISPEQALDAQVDYLRVLLAKRRQPADVYSGLVQADES  
GOLSEAEVLSMAHLMSGFETTMIGNALVTLVNPEDQLARQAPELIPAMEL  
VRHDSVFRASMLRTYEDVELDGYTIPAGEYILVSNLTANHDAERDDPDRLTENT  
DGLIGYGFVHYCVGASLARLEGRITAIORLLARPPDQILAVPAHELQWLPITFLALI  
SVPRPGCSAPANTASHANPIERIAQ"  
12997..13422  
/gene="PA3332"  
12997..13422  
/gene="PA3332"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAG06720.1"  
/db\_xref="GI:9949463"  
/translation="MAKELVHSRLIENGANGKCDLHPPEGVLEFPYAPGKTR  
FEGREITMAHMLFPEHLTVRFTVOQFETVADPLAIGEHGQVATVSGKLADYI  
SVLRTRDGOILLYRDFWNLRLHEALGVEAARIVOGA"  
13426..14418  
/gene="fabH2"  
13426..14418  
/note="PA3333"  
/gene="fabH2"  
/codon\_start=1  
/transl\_table=11  
/product="3-oxoacyl-[acyl-carrier-protein]  
/protein\_id="AAG06721.1"  
/db\_xref="GI:9949464"  
/translation="MPRAVYVGGISYLPKAVLSNDMLAELDPTSAWISSRTGVOR  
HIRGDSGLDILRAASALIASAGLERDAVLTLSNGDCCCATAPRAARIGVGA  
LAPDLNACTGTYTGLASGSLISNGHLDLMTAVYTPFSITLDRDRSRALFEGGA  
GAUVLRAGDAEERGALLAPDUSGHQPDLMTRPAVRAERSGQSNFRMDKAYE  
GOAVTQMSDSVRVLDYRGWQASDLHLVLPQAVRILIAVAQDLDPVERVSNIAE  
VGNTVAAISIPALAHGLRGLIRDGGNNVILNFGAGLWGSVALRMPKIVPTMD"  
14432..14671  
/gene="PA3334"  
14432..14671  
/gene="PA3334"  
/codon\_start=1  
/transl\_table=11  
/product="probable acyl carrier protein"  
Query Match 16.9%; Score 204.4; DB 1; Length 14759;  
Best Local Similarity 53.0%; Pred. No. 1.5e-08;  
Matches 595; Conservative 0; Mismatches 491; Indels 36; Gaps 6;  
QY 84 ggagcaagaccggtgacccggtgagccgacccgaggtctggtctcgcgtccgctcgagga 143  
DB 11827 gccgcgtcccgccgcatgcccccgcgacgcatccgggtggtgagcgcgtacccgga 11886  
QY 144 cgtgcgcgagtgctgcgacagacgcttcgcgcg----gaccggaatccgacggtctc 199  
DB 11887 ccgcgcgcaaggtctcctcaacatccggcgccgcgcgacgcccgcgacgcccgaact 11946  
QY 200 ggaacgactcgt-----cggcgcgaggtcgtcgcgcgagatcgcgcgacgacatc 251  
DB 11947 ctacgcgaagcgtaccgcgaccccgccgaggtccgacgagcactacgcacacat 12006

OY	252	cccttcacgaagagaccgcacacacagccagagctctgcagggatgtatctgcacccgaggtcttc	311
Db	12007	gcttcaacctgcagaccgccgcgacacattacccgccttcgctccttgatggaccgcgggtttcac	12066
OY	312	gcgcgtccgcgtctgcgcgcgacttgaaacaggtatgcgcgcgcacacgtltgaacgaactgtctgc	371
Db	12067	ccccgcgcacagatggagagccctctgcacacccgattatgaacggatcaccocgagcattgcgga	12126
OY	372	gcccccccttgcccgagggcgagatgaagctgtgtcacagagctgcgcgtataccgcctgcgact	431
Db	12127	gcgcattggccggcgccgaacagaccgcgacctgattggcccgactttcgagatcccgctgacatt	12186
OY	432	gcgcgcggtctctgcgcctctgtgccttgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	491
Db	12187	cgccgctattcttgcagcttgccttgggcatttcccgagccgacgcgcgcgcgcgcgcgcgcgc	12238
OY	492	gtgcgcgcagctgtggagcgcgccttgagccggggcgccagcgccgcgcgcgcgcgcgcgcgcgc	551
Db	12239	--gcgcacgtctctggagccgc	12296
OY	552	ccacgc	611
Db	12297	cc-----gattgcgagagctgcacattacctggcgcgtgcgtgcgcgcgcgcgcgcgcgcgcgc	12348
OY	612	tgaaagcgagcgagagacactgtgcgtgtgtatgtctgcagcccaacgcgcgcgcgcgcgcgcgc	671
Db	12349	gccgc	12408
OY	672	tcgcacacgagatcgtcagcaagctgtgtcaactgttcattcttcacgcgcgcgcgcgcgcgcgc	731
Db	12409	cgaaacgcgaacctgcgttccattgc	12468
OY	732	cagcacaagctggagcaacgcgcgcgtgtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	791
Db	12469	gaacatgatggcgcaacgc	12528
OY	792	ccgc	851
Db	12529	gcgggc	12588
OY	852	gggtcagctccaacacccgc	908
Db	12589	ggtcgc	12648
OY	909	gc	968
Db	12649	tcgc	12708
OY	969	gc	1028
Db	12709	gc	12768
OY	1029	gggagctgcctactgcctcgcgtctcctacccctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1088
Db	12769	cgcgctgctacattactgctgc	12828
OY	1089	gcgcctc---gc	1145
Db	12829	gc	12888
OY	1146	tcagcgcgcgcacatgtctccgc	1187
Db	12889	gctccgc	12930
RESULT	12		
ABO18074			
LOCUS	ABO18074	2795 bp	DNA linear BCT 06-OCT-1998
DEFINITION	Streptomyces griseus genes for orf-1 and orf-2, similar to chalcone synthase and cytochrome P-450, complete cds.		
ACCESSION	ABO18074	BA5916	
VERSION	ABO18074.1	GI:3702259	
KEYWORDS	cytochrome P-450; chalcone synthase.		

	SOURCE	Streptomyces griseus DNA.
	ORGANISM	Streptomyces griseus
		Bacteria; Firmicutes; Actinobacteria; Actinobacteriales;
		Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
	REFERENCE	1 (bases 1 to 2795)
	AUTHORS	Ueda, K.
	TITLE	Direct Submission
	JOURNAL	Submitted (03-OCT-1998) Kenji Ueda, The University of Tokyo,
	COMMENT	Faculty of Agriculture, Department of Agricultural Chemistry,
	MEDLINE	Yayoi, Bunkyo-ku, Tokyo 113, Japan (Tel:03-3812-2111(ex.5147),
		Fax:03-3812-0544)
		2 (sites)
	REFERENCES	Ueda,K., Kim,K.M., Beppu,T. and Horinouchi,S. Overexpression of a gene cluster encoding a chalcone synthase-like protein confers reddrown pigment production in Streptomyces griseus
	JOURNAL	J. Antibiot. 48 (7), 638-646 (1995)
	COMMENT	95378086
		On Oct 6, 1998 this sequence version replaced gi:666020.
	FEATURES	D45916:Submitted (13-Feb-1995).
		Location/Qualifiers
	source	1..2795
		/organism="Streptomyces griseus"
		/db_xref="taxon:1911"
	RBS	330..335
	gene	341..1561
		/gene="orf-2"
	CDS	341..1561
		/gene="orf-2"
		/note="similar to bacterial cytochrome P-450"
		/codon_start=1
		/transl_table=1
		/protein_id="BAA33494.1"
		/db_xref="gi:3702260"
		/translation="MENTSYONKENVNCPDYAHLEFPDROLRLPEPVSIRIRM YGGGAMLVTRREDEVRTVTPDRSRSAVLGRPPMPPTPVQASINILAMPASSH LRIIVAKSPTRPVVGOMCGTGVRYDDPLIDMEESSPCDFVARYSAPLTITCEAL DIFEARPMFLRAHAATMMNGVCAGAKODAVRKAABKSTFDELDARRRSFGCDLLSTI AIRDDELDDLDELAVMAVYLITGDOTTTLTGNTATLLTRPLSKSLAEFORA PRLEELHLHPFRKGVIPIRIALEDELSELGIKAGDVHVHSYLIANNDSKKFERP EPLDPDPTIPMHFTFGMAHCLGAPLATMELEVAFFSTLLTRPALRLDVPEDYSWNN TSIMRPDLAPVTW"
	RBS	1585..1590
	gene	1596..2714
		/gene="orf-1"
	CDS	1596..2714
		/gene="orf-1"
		/note="similar to chalcone synthase"
		/codon_start=1
		/transl_table=1
		/protein_id="BAA33495.1"
		/db_xref="gi:3702261"
		/translation="MATLCRAIAVPEHVITMOOTIDLARETHAGHPORDVLRILION TGQTURLIOPIELEKTLAHGFPEVNQVEAEARTRVPEVVRRLANAETESSEIDLIYA LYSCFEFMPSLTAMIINSMGFPRETROQLPAIDCCAGAALINRAHDCEVAYPSDNV LIVSCFCSICYPOTDIDIGVLSNLGFDGLSAVAVRGGGTGNRLBNENSHLYPDPT EMISIAVDRTGEHFOLDKRVPMTGMELAVLIDVLDLGMSVPMNDFFIYHAGPRP LDICHLIDPEMRRYSRATITRGNIASVYFDLARLPDDGAALSAOGLIANGRC PGTTAAVANGSMARKGLADADVGRUDLELELTAGVALSG"
	BASE COUNT	435 a 1051 c 908 g 401 t
	ORIGIN	
	Query Match	16.8%; Score 203.6; DB 1; Length 2795;
	Best Local Similarity	52.1%; Pred. No. 3,4e+08;
	Matches 586; Conservative	0; Mismatches 514; Indels 24; Gaps 5;
Db	448	GCCGCTGCCGACATCCGTAATGCGCATTGCAAGAGGCGAGGCCTGCTGATCACCCGCTA 507
Qy	78	gctcgtcgagcaagaccgcggtatgcacccgtggcgccacccggtcttcgtgaagtcccgctt 137
Db	448	GCCGCTGCCGACATCCGTAATGCGCATTGCAAGAGGCGAGGCCTGCTGATCACCCGCTA 507
Qy	138	cgcgcagctgcgcgcgcggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgtt 197
Db	508	CGAGAGATCCCGACGCGTCACACCACGCGCGGTTCAGCCGACGCGCTCTCGGCGG 567

[illegible][illegible]

Db 919 GGAGCT-----GCGGGCTACTTCCAGGAGCTGACGCCGACGCCGCCGCTCCCGGG 972  
 Qy 618 cggcggagacgtctgagcgttgatctcagcggccacggacggcgtgatagtcgca 677  
 Db 973 CGAGGACCTCATCAGACACCTGGCCACCCCGGAGCGAGCAAGTCTGTGAGACGA 1032  
 Qy 678 cggatcgtcagcaggtgtgtaacgttcaatcttcacggccacgagcgggttgacgca 737  
 Db 1033 CGAGCTGGCGCTATGACCATGTCTGTCTATCACCGGCGAGACACACAGCATACCA 1092  
 Qy 738 ggtgggcaacgcgctgtgagcctgtggcgacccggacgcgctgacgtctccggcg 797  
 Db 1093 GCTCGGCAATGCGCTACACCTGTCTACACCGCCGGACGCTGCTCGGCTCCGGGCG 1152  
 Qy 798 cggcggagacgtctgagcggcggcggcggcggcggcggcggcggcggcggcggc 851  
 Db 1153 CGAAGCGGACGGCTGCGCCCGGACCCCTGGAGAGCTGCTGGCCACATCCCTTCGCA 1212  
 Qy 852 ggtgcagctcaaacaccggcggcggcggcggcggcggcggcggcggcggcggcggc 911  
 Db 1213 GGGCGTGGGATCCCGCTGATCGCCCTGGAGAGCTGCTGCGGCTCCTCATCA 1272  
 Qy 912 cggcggagacgtctgagcgttgctgagcggcggcggcggcggcggcggcggcggc 971  
 Db 1273 GGGCGGAGAGCTGTGACGCTGTCTACCTGACGGCCACCGGACTCCGCCAAGTCA 1332  
 Qy 972 cggcggagacgtctgagcggcggcggcggcggcggcggcggcggcggcggcggc 1031  
 Db 1333 CCGTCCCGAGAGCTGGACCCCGGACCGGCGGACCATATCCCATGCTGGCTGGG 1392  
 Qy 1032 gatgagctactgctggcggtctgaactgcggcgttaagcagctgagcggcggcggc 1091  
 Db 1393 CGGCCACACACTGCTGGGCGCGCGCTGAGCCACCATGAGACCTGGAAGGCTTCTCAC 1452  
 Qy 1092 ---cctggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggc 1148  
 Db 1453 GCTGCTACCCGCTTCCCGGCGCTGCTGAGACGTGCGCCCGGAGAGCTCTCGTGA 1512  
 Qy 1149 gccgcgacacatgttcggggcgtggcggcggcggcggcggcggcggcggcggc 1192  
 Db 1513 CAGACGTCATCTGGCGTTACCGCGTCCGCTGCGCTCATCAT 1556  
 RESULT 14  
 SCE41/c 36028 bp DNA linear BCT 02-NOV-2000  
 LOCUS Streptomyces coelicolor cosmid E41.  
 DEFINITION AL442120  
 ACCESSION AL442120  
 VERSION AL442120.1 GI:10241774  
 KEYWORDS ABC transport system ATP-binding protein; ABC transport system integral membrane protein; cytochrome P450 hydroxylase; eno, enolase; hydrolase; integral membrane protein; lipoprotein; nucleotidyltransferase; oxidoreductase; pkae; secreted protein; sensor kinase; transcriptional-repair coupling factor.  
 SOURCE Streptomyces coelicolor.  
 ORGANISM Streptomyces coelicolor  
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 REFERENCE 1 (bases 1 to 36028)  
 AUTHORS Redenbach,M., Kleser,H.M., Denapalte,D., Eichner,A., Cullum,J., Kinashi,H. and Hopwood,D.A.  
 TITLE A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome  
 JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)  
 MEDLINE 97000351  
 REFERENCE 2 (bases 1 to 36028)  
 AUTHORS Saunders,D.C. and Harris,D.  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 36028)  
 AUTHORS Cerdano,A.M., Parkhill,J., Barrell,B.G. and Rastam, M.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (19-SEP-2000) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge

## COMMENT

CB10 15A E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK  
 Notes:  
 Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics  
 details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.  
 (URL: [http://www.sanger.ac.uk/Projects/S\\_coelicolor/](http://www.sanger.ac.uk/Projects/S_coelicolor/)) CDS are numbered using the following system eg SC/B7.01c, SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).  
 The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.  
 Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, tgg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.  
 IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid E41.

## FEATURES

## source

1..36028

/organism="Streptomyces coelicolor"

/db\_xref="taxon:1902"

complement(1..951)

/gene="SCE41.01c"

complement(1..951)

/gene="SCE41.01c"

/note="SCE41.01c, probable oxidoreductase (fragment), len: >317 aa; similar to SW:DHNA\_ECOLI (EMBL:U00306)

Escherichia coli NADH dehydrogenase (EC 1.6.99.3) Ndb, 433 aa; fasta scores: opt: 361 z-score: 417.8 E(): 9.1e-16;

27.6% identity in 330 aa overlap and to TR:CA92372 (EMBL:AL356612) Streptomyces coelicolor putative NADH

dehydrogenase SCD72A.05, 442 aa; fasta scores: opt: 1185 z-score: 1267.3 E(): 0; 57.0% identity in 321 aa overlap

Contains Pfam match to entry PF00070 pyr\_redox, Pyridine nucleotide-disulphide oxidoreductase"

/codon\_start=1

/transl\_table=1

/product="putative oxidoreductase (fragment)"

/protein\_id="CAC09533.1"

/db\_xref="GI:10241775"

/translation="MRYGEATVTVVDRSRMYTQPLPEAAGSISPRHVVPLRVYL PKAEVLGRVTTIDODRKVAIVAPLVEAELEFPDYIVAMGAVSRFFPGLAEOGI

GKAEIESIGLRNVLFDLPDADSTDEELERKALFVFGAGAGETIGEDMAR DAATYNNVSRKEDRFLIVDAADKILPEVGPRIKGVCKELEGVGVLYSTMDSCV

DGHVAVLNGLDEVDSNTITWAGVAPNPAALRFEGILPEPRRHVDTQATLVQSGDTIYA

AGDNAOVDPDLGRRAGNEMACPNPAHALRQAARLGDVNI"

1..101

/note="nominal overlap with Streptomyces coelicolor cosmid SCE25"

SCE25"

1..36028

/organism="Streptomyces coelicolor A3(2)"

/strain="A3(2)"

/db\_xref="taxon:100226"

/clone="cosmid E41"

complement(112..798)

/gene="SCE41.01c"

## misc-feature





Query Match	Best Local Similarity	Score	DB 1:	Length
Matches 592; Conservative 0; Mismatches 479; Indels 39; Gaps 7;	53.3%;	195.6;	1.9e+07;	1400;
Query 98	tgcacgctggcgcccaacgggtctgtgtacgtctcccttcgagagcttgccgagctgc	157		
Db 212	tcacctacggcgagggcagcggcctgtgctgtacccgcatgctccagcccgatctgttc	301		
Qy 158	tcgagcagcagcgtctgcgcccgagccgagcatccgcggtcttgacccagctctgcgagc	217		
Db 302	tggcgagactcccgcttccagcagccggcgccgacccgacccgaccccgccgaccccgat	356		
Qy 218	ccggagctgcgcgcgagatcgttcgagacatcatctcgttccagagcagcccgacacag	277		
Db 357	---gttccccaccgccggccgagccgagccgagccgctgctccagcagcccgccgacaca	412		
Qy 278	gccgctcgcgcgggtgtgttcggccgggtcttcgcgctcgcgctgcgcgctgcgcgagac	337		
Db 413	ccggctcgcgcgggtgtgttcggccgggtcttcgcgctcgcgctgcgcgctgcgcgagac	472		
Qy 338	cgggtgc	397		
Db 473	ccggctcgc	532		
Qy 398	acgtgtgc	454		
Db 533	ccgacccgctgc	592		
Qy 455	gctcgc	514		
Db 593	gctcgc	649		
Qy 515	tgagccggcgagccgc	574		
Db 650	cgacccggcgagccgc	697		
Qy 575	ccgactacgtgagcgagccctcgc	632		
Db 698	tggctacatgagc	757		
Qy 633	ggcgtgtgagtgagc	691		
Db 758	ggcgtgtgagtgagc	817		
Qy 692	cgggtgtgagtgagc	751		
Db 818	tgggggtgagtgagc	877		
Qy 752	tgctgagctgcgtgc	811		
Db 878	tcacacctcgtgc	937		
Qy 812	tgcccgagcgctgcagagtgctgcgtgcgtacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	865		

```

Db 938 TGCCCGGCGGCGTGAGAGATGCTGCGGTACACACCCGTGTTGCCCGGACAGTTTCG 997
QY 866 cccgcagctcgaagctgaagctgagctgagctgagctgagctgagctgagctgagctg 925
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 998 TCCCGGTGGCCACCGAGGACCTGAGCTGAGCACCCGTGACCGTTCGGGGCCGGGAGCCCT 1057
QY 926 tggctcctcggccgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 985
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1058 GCGTCGTCCACTTCGCGTCGGCCAAACCGGGACGAGAGTCTTCGACACGCGCAGCAGC 1117
QY 986 tcgacatcgaagcgagatccggtccgctcgatgcttcgagcgagcgagcgagcgagc 1045
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1118 TGGACTTCCACCGTGAAGCGCAACCCGACATAGCGTTCCGGGCAACGACGACACCTGCA 1177
QY 1046 tcggtccctacctgcgcgtacgagctgagcgagcgagcgagcgagcgagcgagc--- 1102
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1178 TCGGCGGCCCAACTGGGCGCGACTGGAACCTCAGAGGCCCTTCTCCGCCCTGTCGGGCT 1237
QY 1103 tgcgagcgctgagctgagctgagctgagcgagcgagcgagcgagcgagcgagcgagc 1162
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1238 TCCCGACCTTCGATCTGAGCGGAGCCGCTCGCGGACTGGAAGTGAAGCAGGCGCATGCTGA 1297
QY 1163 tcgagggcgctggcgagcgctgagcgagcgagcgagcgagcgagcgagcgagcgagc 1192
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1298 TCCGGGACTGGAAGCGCCAGATCGTCTCT 1327

```

Search completed: June 21, 2002, 15:52:22  
 Job time: 7353 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2002, 15:13:00 ; Search time 207.34 Seconds  
(Without alignments)  
10011.332 Million cell updates/sec

Title: US-09-724-797-35

Perfect score: 1209  
Sequence: 1 gtcgtgcgtcgtacgtcgtgac.....cgttcacgcgcgcgttga 1209

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_032802:\*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	214.4	17.7	1233	21	AAC55790 Hydroxylase orf3
2	214.4	17.7	12249	21	AAC55840 Complete Mitomycin
3	214.4	17.7	18331	21	AAC55857 Complete nucleotid
4	212.2	17.6	6085	18	AAT70153 S.longisporoflavus
5	204.4	16.9	1257	22	AAAF81357 Quorum sensing con
6	203.6	16.8	2795	22	AAAF81296 Streptomycetes grise
7	195.6	16.2	50937	21	AAAF81954 Streptomycetes grise
8	195.6	16.2	50937	21	AAAF81954 Streptomycetes grise
9	194.8	16.1	21185	21	AAA63350 Streptomycetes globi

10	194.8	16.1	63164	21	AAA63348 Streptomycetes globi
11	194.6	16.1	1224	21	AAC55788 Hydroxylase orf4
12	193.8	16.0	1700	14	AAQ45569 Sequence soyC and
13	187.6	15.5	1400	12	AAQ14548 Eryf gene. Saccar
14	180.8	15.0	4403765	22	AA1996683 Mycobacterium tube
15	179	14.8	5970	21	AAAF6535 Nucleotide sequenc
16	179	14.8	5970	21	AAAF6535 Nucleotide sequenc
17	175.8	14.5	1251	21	AAZ56003 Contig 002 from co
18	169.4	14.0	1998	12	AAQ11127 S. venezuelae macr
19	160.8	13.3	1188	21	AAC55828 Sequence encoding
20	160.8	13.3	53500	21	AAC55842 S. lavendulae MmcN
21	160.2	13.3	2168	15	AAQ73674 Complete nucleotid
22	157.4	13.0	1879	12	AAQ11126 Mycinamicin IV hyd
23	155.4	12.8	47981	22	AAF30757 Sequence encoding
24	155.2	12.8	53789	19	AAV21187 Micromonospora meg
25	152.4	12.6	2918	21	AAZ43922 Amycolatopsis medi
26	152.4	12.6	2919	21	AAZ24486 S. tendae mikkomyc
27	150.4	12.4	5821	18	AAT58686 Streptomycetes tende
28	148.8	12.3	14806	22	AAQ03809 DNA encoding S. fr
29	148.8	12.3	77536	21	AAAI4651 Streptomycetes galli
30	146.8	12.1	1197	22	AAAF2894 Nucleotide sequenc
31	146.6	12.1	123401	22	AAAD17186 Pimaricin biosynth
32	145.4	12.0	12905	21	AAZ99101 Streptomycetes nous
33	143.8	11.9	12905	21	AAZ99100 S. fradiae tylosin
34	142.4	11.8	1194	17	AAT58555 Streptomycetes prist
35	142.4	11.8	4496	17	AAT58553 Streptomycetes prist
36	137.4	11.4	1227	22	AAAD1511 Streptomycetes clavu
37	137.4	11.4	15079	22	AAAD1499 Streptomycetes clavu
38	136.4	11.3	1194	22	AAAF24895 Pimaricin biosynth
39	135.8	11.2	15079	16	AAQ91580 S. clavuligerus cl
40	132.2	10.9	125401	22	AAAD17186 Streptomycetes nous
41	126.6	10.5	8169	19	AAV26609 Actinomadura hibis
42	126	10.4	1233	15	AAQ61452 Cytochrome P450 SC
43	121.6	10.1	12441	21	AAZ87284 S. venezuelae deso
44	121.6	10.1	13613	21	AAZ87319 S. venezuelae deso
45	121.6	10.1	38506	21	AAAF5633 Nucleotide sequenc

#### ALIGNMENTS

RESULT 1	
ID AAC55790	standard: DNA; 1233 BP.
AC AAC5790;	
DT 19-JAN-2001	(first entry)
DE Hydroxylase orf3 involved in MC biosynthesis.	
KW Mitomycin; biosynthesis; mitosome ring system; antibiotic; anti-cancer;	
KW anti-inflammatory; immune-enhancer; immunosuppressant; asthma;	
KW chronic obstructive pulmonary disease; respiratory inflammation;	
KW fungicide; pesticide; ds.	
OS Streptomycetes lavendulae.	
PM WO200053737-A2.	
PD 14-SEP-2000.	
PF 10-MAR-2000; 2000WO-US06394.	
PR 12-MAR-1999; 99US-0266965.	
PA (MIND) UNIV MINNESOTA.	
PA (SHER) SHERMAN D H.	
PA (MAOY) MAO Y.	
PA (VARO) VAROGLU M.	
PA (HEMM) HE M.	
PA (SHEL) SHELTON P C.	









[illegible]

Query Match	16.9%	Score 204.4	DB 22	Length 1257
Best local Similarity	53.0%	Pred. No. 6.5e-20		
Mismatches 595; Conservative	0	Mismatches 491; Indels	36	Gaps 6

[illegible]





[illegible]





FT		/tag- t	"O-acyl transferase"
FT	CDS	/product=	(24986..25564)
FT		/tag- u	"epimerase"
FT	CDS	/product=	(25815..27170)
FT		/tag- v	"product= "monooxygenase"
FT	CDS	/tag- w	(27214..28593)
FT		/product=	"glycosyl transferase"
FT	CDS	/tag- x	(28590..29588)
FT		/product=	"GDP glucose dehydratase"
FT	CDS	/tag- y	(29632..31197)
FT		/product=	"Scb transmembrane efflux protein"
FT	CDS	/tag- z	(31280..32590)
FT		/tag- z	"tag- z
FT	CDS	/product=	"coenzyme F390 synthetase"
FT		/tag- aa	(32809..34392)
FT	CDS	/product=	"hydroxylase"
FT		/tag- ab	(34458..35294)
FT	CDS	/product=	"ORF 22 protein"
FT		/transl_except=	(pos:35223..35225,aa:Ile)
FT		/transl_except=	(pos:35226..35228,aa:Ala)
FT	CDS	/tag- ac	"complement (35518..35938)"
FT		/product=	"ORF 18 protein"
FT	CDS	/tag- ad	(36165..37490)
FT		/product=	"coenzyme F390 synthetase"
FT	CDS	/tag- ae	(37559..38938)
FT		/product=	"type II NRPS condensation enzyme"
FT	CDS	/tag- af	(38983..39264)
FT		/product=	"type II peptidyl carrier protein"
FT	CDS	/tag- ag	"complement (39367..40986)"
FT		/product=	"aminomutase"
FT	CDS	/tag- ah	"complement (41052..42611)"
FT		/product=	"type II NRPS adenylation enzyme"
FT	CDS	/tag- ai	(43345..446024)
FT		/product=	"transmembrane transport protein"
FT	CDS	/tag- aj	(46167..47171)
FT		/product=	"O-methyl transferase"
FT	CDS	/tag- ak	(47228..48484)
FT		/product=	"P450 hydroxylase"
FT	CDS	/tag- al	"complement (48607..49714)"
FT		/product=	"oxidoreductase"
FT	CDS	/tag- am	(50350..51390)
FT		/product=	"ORF 31 protein"
FT	CDS	/tag- an	"complement (51421..52341)"
FT		/product=	"oxidoeductase"
FT	CDS	/tag- ao	(53241..54074)
FT		/product=	"ORF 33 protein"
FT	CDS	/tag- ap	(54231..55379)
FT		/product=	"ORF 34 protein"
FT	CDS	/tag- aq	(56026..56880)
FT		/product=	"proline oxidase"

FT	CDS	/product= "P450 hydroxylase"
FT		/tag= ar
FT		/product= "ORF 36 protein"
FT	CDS	complement (57833..58304)
FT		/tag= as
FT		/product= "ORF 37 protein"
FT	CDS	complement (58440..60095)
FT		/tag= at
FT		/product= "CDS"
FT	CDS	complement (60092..60621)
FT		/tag= au
FT		/product= "ORF 39 protein"
FT		60940..62019
FT		*tag= av
FT		/product= "ORF 40 protein"
FT	CDS	62045..62899
FT		*tag= aw
FT		/product= "ORF 41 protein"
FT	CDS	complement (62787..63164)
FT		*tag= ax
FT		/product= "ORF 42 protein"
PX		
PX	WO200040596-A1.	
PX		
PX	13-JUL-2000.	
PX		
PF	06-JAN-2000; 2000MO-US00446.	
PR		
PR	06-JAN-1999; 99US-0115434.	
PR	05-JAN-2000; 2000US-0477962.	
PA	(REGC ) UNITV CALIFORNIA.	
PI		
PI	Shen B, Liu W, Christenson SD, Standage S;	
XZ		
XX	WPI: 2000-465947/40.	
DR	P-FSDB: AAB13554, AAB13555, AAB13556, AAB13557, AAB13558, AAB13559,	
DR	AAB13560, AAB13561, AAB13562, AAB13563, AAB13564, AAB13565, AAB13566,	
DR	AAB13567, AAB13568, AAB13569, AAB13570, AAB13571, AAB13572, AAB13573,	
DR	AAB13574, AAB13575, AAB13576, AAB13577, AAB13578, AAB13579, AAB13580,	
DR	AAB13581, AAB13582, AAB13583, AAB13584, AAB13585, AAB13586, AAB13587,	
DR	AAB13588, AAB13589, AAB13590, AAB13591, AAB13592, AAB13593, AAB13594,	
DR	AAB13595, AAB13596, AAB13597, AAB13598, AAB13600, AAB13601, AAB13602, AAB13603,	
PT		
PT	Isolated nucleic acid comprising a nucleic acid encoding any of C-1027	
PT	open reading frames (ORFs) -7 to 42, excluding ORF 9 (caga), useful for	
PT	the production of enediyne C-1027 antitumour antibiotics -	
PS	Claim 1: Page 78-157; 160pp; English.	
CC	The present sequence is the enediyne C-1027 gene cluster from	
CC	Streptomyces globosporus. Enediyne C-1027 is an antibiotic, consisting of	
CC	an apoprotein and a non-peptide chromophore, which acts by damaging DNA.	
CC	The sequences within the gene cluster, and the proteins they encode, can	
CC	be used in the treatment of cancer, along with antagonists of the	
QY	Query Match	16.1%; Score 194.8; DB 21; Length 63164;
QY	Best Local Similarity	51.8%; Pred. No. 6.5e-19;
QY	Matches 604; Conservative	0; Mismatches 532; Indels 30; Gaps 6;
DG	44 acgtgcgcttgaccctacccgttcactaccatgctctcgagcaac--aacccagtgc 100   Dg 47283 agggcgctcgaaggccaccttacccttctaaggagcatcgcgcgacggcgagctgcc 47342	
OY	101 accgttgtgcgccacccgggtcttgttaagctctcccgcgttcgcggagcgtgcggcggtgctcg 160   Oy 47343 ccacgaagccccgcgcgctctgtgtctgcgcacgaagcgcgtgcgacggcgctgcgcgg 47402	
OY	161 ggcagcacggcgttcggccgacgcgcgcgcgc--cgattctgaacacgcacgtctgcggcc 218   Db 47403 aaccggcctcgcgcgttcataagacgacgagggcctaaagcgcgcgtcttcctcctgcggcggt 47462	
OY	219 cgggctgcctc-gccgagatcgttcgcgacatatcctctgtttccaagacgagcccgacaacg 277	



[illegible]

RESULT	11
AAC55788	
ID	AAC55788 standard; DNA; 1224 BP.
XX	
AC	AAC55788:

XX 19-JAN-2001 (first entry)  
DT Hydroxylase orf4 involved in MC biosynthesis.  
XX  
XX  
XX Mitomycin: biosynthesis; mitosome ring system; antibiotic; anti-cancer;  
KM anti-inflammatory; immune-enhancer; immunosuppressant; asthma;  
KW chronic obstructive pulmonary disease; respiratory inflammation;  
XX fungicide; pesticide; ds.  
XX  
OS Streptomyces lavendulae.  
XX  
PN WO200053737-A2.  
PD  
PD 14-SEP-2000.  
XX  
XX  
PF 10-MAR-2000; 2000WO-US06394.  
PR  
PR 12-MAR-1999; 99US-0266965.  
XX  
XX (MIND ) UNIV MINNESOTA.  
PA (SHER/) SHERMAN D H.  
PA (MAOY/) MAO Y.  
PA (VARO/) VAROGLU M.  
PA (HEM/) HE M.  
PA (SHEL/) SHELDON P C.  
XX  
PI Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC.  
PI  
DR WPI: 2000-601980/57.  
PT  
PT Novel nucleic acid molecule comprising mitomycin biosynthetic gene  
PT cluster useful for cloning mitomycin biosynthetic genes for elucidating  
PT the molecular basis of mitosome ring system biosynthesis -  
XX  
XX  
PS Example 1: Page 254; 39pp; English.  
XX  
XX This invention relates to isolated and purified nucleic acid molecules  
CC from the mitomycin biosynthetic gene cluster. Mitomycins are a group of  
CC natural products that contain a variety of functional groups, including  
CC amino benzquinone and axiridine ring systems. The S. lavendulae  
CC mitomycin biosynthetic gene cluster comprises 47 mitomycin genes  
CC spanning 55kb of DNA. The invention includes an expression cassette  
CC comprising a mitomycin biosynthetic gene operably linked to a promoter,  
CC and host cells transformed with the cassette. The nucleotide, and protein  
CC sequences and the transformed host cells of the invention result in  
CC antiasthmatic, antiinflammatory, cytostatic, immunomodulatory, and  
CC antibiotic activities. The nucleotide sequences are used to elucidate the  
CC molecular basis for the biosynthesis of the mitosome ring system, as well  
CC as to engineer the biosynthesis of novel natural products, e.g.  
CC antibiotics, anti-inflammatory agents, anti-cancer agents,  
CC immune-enhancers, immunosuppressants, agents to treat asthma, chronic  
CC obstructive pulmonary disease as well as other disease involving  
CC respiratory inflammation, or cholesterol-lowering agents or as crop  
CC protection agents (e.g. fungicides or insecticides) as well as  
CC biopolymers, e.g., in packaging or biomedical applications, or to engineer  
CC PAA monomer synthases. Sequences AAC55782-C55881, AAC5815-C55849 and  
CC AAB32485-B35542 represent mitomycin biosynthetic gene cluster DNA  
CC sequences and encoded proteins. Sequences AAC55812-C55814,  
CC AAC55850-C55856 and AAC55862-C55869 represent PCR primers used in the  
CC cloning of the mitomycin biosynthetic genes.  
XX  
XX Sequence 1224 BP; 196 A; 492 C; 379 G; 157 T; 0 other;

[illegible]



[illegible][illegible]

[illegible][illegible]





```

RESULT 1
US-09-029-603-4      : Sequence 4, Application US/09029603
                        : Patent No. 6210935
                        :
GENERAL INFORMATION:
  APPLICANT: Schupp, Thomas
  APPLICANT: Engel, Natalie
  APPLICANT: Bietenhader, Jurg
  APPLICANT: Toupet, Christine
  APPLICANT: Pospishech, Andreas
  TITLE OF INVENTION: Staurosporin Biosynthesis Gene Clusters
  FILE REFERENCE: 4-2055/A/PCT
  CURRENT APPLICATION NUMBER: US/09/029, 603
  CURRENT FILING DATE: 1998-03-20
  EARLIER APPLICATION NUMBER: PCT/EP96/03643
  EARLIER FILING DATE: 1996-08-19
  NUMBER OF SEQ ID NOS: 11
  SOFTWARE: PatentIn Ver. 2.0
  SEQ ID NO 4
  LENGTH: 6085
  TYPE: DNA
  ORGANISM: Streptomyces longisporoflavus
  FEATURE:
  NAME/KEY: misc_RNA
  LOCATION: (378)..(1665)
  OTHER INFORMATION: ORF
  FEATURE:
  NAME/KEY: misc_RNA
  LOCATION: (1747)..(2553)
  OTHER INFORMATION: ORF
  FEATURE:
  NAME/KEY: misc_RNA
  LOCATION: (2593)..(4011)
  OTHER INFORMATION: ORF
  FEATURE:
  NAME/KEY: misc_RNA
  LOCATION: (4013)..(4999)
  OTHER INFORMATION: ORF
  FEATURE:
  NAME/KEY: misc_RNA
  LOCATION: (5071)..(6085)
  OTHER INFORMATION: ORF
US-09-029-603-4

Query Match      17.6%; Score 212.2; DB 4; Length 6085;
Best Local Similarity 52.0%; Pred. No. 1.3e-26;
Matches 555; Conservative 0; Mismatches 478; Indels 6; Gaps 2
90 cgaccgcgtgacacgctgacgccacacggtctcgtgacgtctccgcgcttcgcgagctgacg 149
yy

```





[illegible]

3  
 US-08-102-863-10  
 Sequence 10, Application US/08102863  
 Patent No. 3465390  
 GENERAL INFORMATION:  
 APPLICANT: SARISLANI, SIMA  
 TITLE OF INVENTION: CONSTITUTIVE  
 TITLE OF INVENTION: EXPRESSION OF P450SOY  
 TITLE OF INVENTION: AND FERREDOXIN-SOY IN  
 TITLE OF INVENTION: STREPTOMICES  
 NUMBER OF SEQUENCES: 12  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: E. I. DU PONT DE NEMOURS  
 ADDRESSEE: AND COMPANY  
 STREET: 1007 MARKET STREET  
 CITY: WILMINGTON  
 STATE: DELAWARE  
 COUNTRY: USA  
 ZIP: 19898  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0,  
 SOFTWARE: Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/102,863  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/07/807,001  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: GALLEGOS, R. THOMAS  
 REGISTRATION NUMBER: 32,692  
 REFERENCE/DOCKET NUMBER: CR-9000  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 302-892-7342  
 TELEFAX: 302-892-7949  
 INFORMATION FOR SEQ ID NO: 10:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1735 base pairs

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-102-863-10

```

Query Match	16.0%;	Score 193.8;	DB 1;	Length 1735;
Best Local Similarity	51.4%;	Pred. No. 1.1e-23;		
Matches 562;	Conservative	0;	Mismatches 507;	Indels 24;
				Gaps 4;

[illegible]

Accession	Sequence	Position
Db 1225	GCCCGCATACCTCGGCTTCGGCTTCGGAGTCCACAGTGGCTTG6GCCAAGACTGTGCC	1284
Qy 1063	cgtaacagactgtgcgcgcgcgcgtgtgcgcgcgtg--gcccagactgcgcgcgcgtgtgtgtg	1119
Db 1285	CGCGCCGAGCTGGGACATCCGGGATGGGCACCTGTTTGAGCGGGCTTCCGGGGCTCAGGCTC	1344
Qy 1120	ggtctgcgcgttcgcgcgcctgtgcctatcagcgcgcgcacatgtttccgcgggcctgtgcacg	1179
Db 1345	GCCCTACACCCCGACGACGATCCCTCACAAGCCGGGGAGACACATCCAGAGGCTCTCTCGAC	1404
Qy 1180	ctgcgcagtcgcgt 1192	
Db 1405	CTGGCCGTTGGGCT 1417	

```

1 RESULT 4
2 PCT-US92-10885-10
3 Sequence 10, Application PC/TUS9210885
4 GENERAL INFORMATION:
5 APPLICANT: SARISLANI, SIMA
6 TITLE OF INVENTION: CONSTITUTIVE
7 TITLE OF INVENTION: EXPRESSION OF P450SOY
8 TITLE OF INVENTION: AND FERREDOXIN-SOY IN
9 TITLE OF INVENTION: STREPTOMYCES
10 NUMBER OF SEQUENCES: 11
11 CORRESPONDENCE ADDRESS:
12 ADDRESSEE: E. I. DU PONT DE NEMOURS
13 ADDRESSEE: AND COMPANY
14 STREET: 1007 MARKET STREET
15 CITY: WILMINGTON
16 STATE: DELAWARE
17 COUNTRY: USA
18 ZIP: 19898
19 COMPUTER READABLE FORM:
20 MEDIUM TYPE: Diskette, 3.50 inch,
21 MEDIUM TYPE: 1.0 MB
22 COMPUTER: MacIntosh
23 OPERATING SYSTEM: Macintosh System, 6.0
24 SOFTWARE: Microsoft Word, 4.0
25 CURRENT APPLICATION DATA:
26 APPLICATION NUMBER: PCT/US92/10885
27 FILING DATE: 19921216
28 CLASSIFICATION:
29 ATTORNEY/AGENT INFORMATION:
30 NAME: GALLEGOS, R. THOMAS
31 REGISTRATION NUMBER: 32,692
32 REFERENCE/DOCKET NUMBER: CR-9000-A
33 TELECOMMUNICATION INFORMATION:
34 TELEPHONE: 302-892-7342
35 TELEFAX: 302-892-7949
36 INFORMATION FOR SEQ ID NO: 10:
37 SEQUENCE CHARACTERISTICS:
38 LENGTH: 1735 base pairs
39 TYPE: NUCLEIC ACID
40 STRANDEDNESS: Single
41 TOPOLOGY: linear
42 MOLECULE TYPE: DNA (genomic)
43 PCT-US92-10885-10

```

Query Match	16.0%;	Score 193.8;	DB 5;	Length 1735;
Best Local Similarity	51.4%;	Pred. No. 1.1e-23;		
Matches 562; Conservative	0;	Mismatches 507;	Indels 24;	Gaps 4;

QY 106 gggagcccccggggtctgttaagttctccggtctcgcgagactgtgcggggtcgcggcgcgc 165  
Db 343 GAGCGAGACGCGCGGTGTGAGGCGGTATACCGGAGACGCCCTGGCCCGTGGCTACTGGCGGAC 402  
QY 166 gagcgcttcgcgcgcgacgcgcatccgcgcggttcgtgaacgacctcgttcgggccgggctg 225  
Db 403 CCGCGGCTCTCCACGACCGAGCGCACCCGGACTTCCCCGTCGCCGGCGAGCGGCTTCGCC 462  
QY 226 ctgcgcgagatcgttcgcgcgacatcatcctgttccagagacgagccgcgacacgcgcgctg 285

[illegible]

RESULT 5  
US-09-103-840A-2/c  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:

RESULT 5  
US-09-103-840A-2/C  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:







```
;TEPPERMAN, JAMES M.
; TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
; CYTOCHROMES
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/569,781
; FILING DATE: 23-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 464,499
; FILING DATE: 12-JAN-1990
; APPLICATION NUMBER: 405,605
; FILING DATE: 11-SEP-1989
; SEQ ID NO:8
; LENGTH: 1998
5212296-8
```

```
Query Match 14.0%; Score 169.4; DB 6; Length 1998;
Best Local Similarity 50.3%; Pred. No. 9.1e-20;
Matches 611; Conservative 0; Mismatches 556; Indels 48; Gaps 6;
```

```
QY 9 cgatgcgctgacgcgctgctgataccaccgacgcgcgagctgacgcgctgacccctacccgctc 68
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 224 cgagccctacacgctcccgccagccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcga 283
QY 69 ctaccactgctgctgacgacgacccgctgacacgctgacgcgcgcgcgcgcgcgcgcgcgcgc 118
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 284 cgtaccgagcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 343
QY 119 --ctgtgtagctcccgctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 176
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 344 ctgctgctgctgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 403
QY 177 ccgagccgacatccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 236
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 404 cgccgagccacccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 453
QY 237 cgttcgacgacatcatcctgttccagacgacgacgacgacgacgacgacgacgacgacgacg 296
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 464 caccacccgacccctcctcgtcgtacgacgacgacgacgacgacgacgacgacgacgacgac 523
QY 297 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 356
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 524 caccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 583
QY 357 ggaacgacactgc--gagccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 413
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 584 cgagacgacactgctgcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 643
QY 414 ggcgtaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 473
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 644 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 703
QY 474 ggggggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 533
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 704 cgcgtctctccagagcgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 754
QY 534 cgaagacatgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 593
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 755 cgagagagctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 814
QY 594 ccttcgacgagcgcgcgctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 653
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 815 caagcgagagcgcgcgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 866
QY 654 cgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 713
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 857 cgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 916
QY 714 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 773
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 917 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 976
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
QY 774 ggaccagctcgcacgtctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 833
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 977 cgaccaactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1036
QY 834 ccttcgagtaacacccgctgctgacgtacgtacgtacgtacgtacgtacgtacgtacgtacgtac 893
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1037 gctgcgtacacgtacgtacgtacgtacgtacgtacgtacgtacgtacgtacgtacgtacgtac 1096
QY 894 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 950
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1097 ccttcgagtaacacccgctgctgacgtacgtacgtacgtacgtacgtacgtacgtacgtac 1156
QY 951 ccggagacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1010
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1157 ccggagacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1216
QY 1011 gtcgagtgctcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1070
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1217 ccagctgctcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1276
QY 1071 gctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1127
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1277 gctccagatcgcacatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1336
QY 1128 gtcgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1187
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1337 ccacgagagatccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1396
QY 1188 cgcgttcaagccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1202
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1397 cgcctgtagccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1411
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 10
5212296-5
; Patent No. 5212296
; APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LIETO, KENNETH
; J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSER, JAMES A.
; TEPPERMAN, JAMES M.
; TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
; CYTOCHROMES
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/569,781
; FILING DATE: 23-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 464,499
; FILING DATE: 12-JAN-1990
; APPLICATION NUMBER: 405,605
; FILING DATE: 11-SEP-1989
; SEQ ID NO:5;
; LENGTH: 1879
5212296-5
```

```
Query Match 13.0%; Score 157.4; DB 6; Length 1879;
Best Local Similarity 49.6%; Pred. No. 7.6e-18;
Matches 559; Conservative 0; Mismatches 541; Indels 27; Gaps 5;
```

```
QY 85 cgacagacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 144
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 245 ctgcacccggtgtagcgtctacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 304
QY 145 gtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 204
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 305 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 364
QY 205 gacccgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 264
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 365 gccacgtaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 424
QY 265 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 324
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```









```
QY 99 gcaacgctggcggccacgggtctgtgacgtctccgcttcgcggaacgtgcgcggtgct 158
   || || || || || || || || || || || || || || || || || || || || ||
Db 129 gctgtggagacggcagccagatgtgctgtgacacctgcacgcccggggccgctct 188
   || || || || || || || || || || || || || || || || || || || || ||
QY 159 cggcgaagcagcgttcgc-----cgacggcgtacgcgcggtctctgacccgactgt 212
   || || || || || || || || || || || || || || || || || || || || ||
Db 189 gggccgaccccgcttcacccgggtgacgagccggccgcttcctccgattcaccggccac 248
   || || || || || || || || || || || || || || || || || || || || ||
QY 213 cgggcccggggtgtcgcgagatcgtgcgacatacatccctgttcacgaagagccga 272
   || || || || || || || || || || || || || || || || || || || || ||
Db 249 ctcccaactgtgtcggcccaacccggagtcggctgttcattccgattgacgacccga 308
   || || || || || || || || || || || || || || || || || || || || ||
QY 273 ccacggcgcggtgtgcgggggtgtgcgcgcgcgcgttcctcgcgttcgcggtgcgcgcgc 331
   || || || || || || || || || || || || || || || || || || || || ||
Db 309 gcactcccgctgcgttcgattgctacccgggacttcctggcccgccggccgagccgct 368
   || || || || || || || || || || || || || || || || || || || || ||
QY 332 tggaaacggtgtatcgcgcggaac--gtgaaacgactgtgcgcgcgcgcgttcgcgcgcgc 389
   || || || || || || || || || || || || || || || || || || || || ||
Db 369 ggcggccggcggttcgggagctgctgacgagatcctggcgggctgtgtgaaaggggagcgc 428
   || || || || || || || || || || || || || || || || || || || || ||
QY 390 cggagatgagcgtgtgtcgaagatgtgcgtacccgcgtgcgcgcgcgttcgcgcgcgc 449
   || || || || || || || || || || || || || || || || || || || || ||
Db 429 gccggtgcgactgtgcgcgagctgacgagatcccggtgcgcgcgcgttcgacacccgtgct 488
   || || || || || || || || || || || || || || || || || || || || ||
QY 450 gctcgcgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 509
   || || || || || || || || || || || || || || || || || || || || ||
Db 489 cttgcgc-----gcccgtgacgaccccggaattcattgagagacgcg 530
   || || || || || || || || || || || || || || || || || || || || ||
QY 510 gacccttgacgggggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 569
   || || || || || || || || || || || || || || || || || || || || ||
Db 531 cagcggggttcctcattgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 590
   || || || || || || || || || || || || || || || || || || || || ||
QY 570 gttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 629
   || || || || || || || || || || || || || || || || || || || || ||
Db 591 actgcaggggtatctgaggagagctgctcagagacgagatcgaacacccggcaccgact 650
   || || || || || || || || || || || || || || || || || || || || ||
QY 630 gctcgcgtgtatgtcgaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 686
   || || || || || || || || || || || || || || || || || || || || ||
Db 651 gatcagcggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 710
   || || || || || || || || || || || || || || || || || || || || ||
QY 687 gagaacggtgtgtacgttcatcttcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 746
   || || || || || || || || || || || || || || || || || || || || ||
Db 711 ccgagatgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 770
   || || || || || || || || || || || || || || || || || || || || ||
QY 747 cgcgcgtgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 806
   || || || || || || || || || || || || || || || || || || || || ||
Db 771 gaggcttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 830
   || || || || || || || || || || || || || || || || || || || || ||
QY 807 cctgtcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 863
   || || || || || || || || || || || || || || || || || || || || ||
Db 831 cctgtcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 890
   || || || || || || || || || || || || || || || || || || || || ||
QY 864 caccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 923
   || || || || || || || || || || || || || || || || || || || || ||
Db 891 gggccggcgccggtgagagacgttcacgctcagatgtcctatccggcgccggcgaggg 950
   || || || || || || || || || || || || || || || || || || || || ||
QY 924 cgtgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 983
   || || || || || || || || || || || || || || || || || || || || ||
Db 951 cgtgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1010
   || || || || || || || || || || || || || || || || || || || || ||
QY 984 ttccagatcgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1043
   || || || || || || || || || || || || || || || || || || || || ||
Db 1011 ggtggagcgtgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1070
   || || || || || || || || || || || || || || || || || || || || ||
QY 1044 cctcgtgtcctactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1100
   || || || || || || || || || || || || || || || || || || || || ||
Db 1071 ccttgccagcgtgctgcccgggtgagcgtgagagatcctcgcgcgcgcgcgcgcgcgcgcgcgcgc 1127
   || || || || || || || || || || || || || || || || || || || || ||
```

```
RESULT 15
US-09-385-028-1
; Sequence 1, Application US/09385028
```

```
; Patent No. 6232106
; GENERAL INFORMATION:
; APPLICANT: Susan E. Jensen
; APPLICANT: Kwamena A. Aidoo
; APPLICANT: Ashish S. Paraskar
; TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
; Patent No. 6232106
; TITLE OF INVENTION: Acid Biosynthesis
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC
; STREET: The Jennifer Building, 400 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/385,028
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/790,462
; FILING DATE: 29-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: D. Douglas Price
; REGISTRATION NUMBER: 24,514
; REFERENCE/DOCKET NUMBER: 1418/P574520S2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 638-6666
; TELEFAX: (202) 39305350
; TELEX: RCA 248593 IDEA UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15079 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces clavuligerus
; US-09-385-028-1

Query Match 11.4%; Score 137.4; DB 4; Length 15079;
Best Local Similarity 50.4%; Pred. No. 1.1e-14;
Matches 513; Conservative 0; Mismatches 471; Indels 33; Gaps 6;

QY 99 gcaacgctggcggccacgggtctgtgacgtctccgcttcgcggaacgtgcgcggtgct 158
   || || || || || || || || || || || || || || || || || || || || ||
Db 13891 gctgtggagacggcagccagatgtgctgtgacacctgcgacccggcgccgctct 13956
   || || || || || || || || || || || || || || || || || || || || ||
QY 159 cggcgaagcagcgttcgc-----cgacggcgtacgcgcggtctctgacccgactgt 212
   || || || || || || || || || || || || || || || || || || || || ||
Db 13957 gggcgacccggcgttcacccggcggtgacgagccggccgcttcctccgattcaccggccac 14016
   || || || || || || || || || || || || || || || || || || || || ||
QY 213 cgggcccggggtgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 272
   || || || || || || || || || || || || || || || || || || || || ||
Db 14017 ctcccaactgtgtcggcccaacccggagtcggctgttcattccgattgacgacccgca 14076
   || || || || || || || || || || || || || || || || || || || || ||
QY 273 ccacggcgcggtgtgcgggggtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 331
   || || || || || || || || || || || || || || || || || || || || ||
Db 14077 gcactcccgctgcgttcgattgctacccgggacttcctgtggcccgccggcggagccct 14136
   || || || || || || || || || || || || || || || || || || || || ||
QY 332 tggaaacggtgtatcgcgcggaac--gtgaaacgactgtgcgcgcgcgcgttcgcgcgcgc 389
   || || || || || || || || || || || || || || || || || || || || ||
Db 14137 ggcggccggcggttcgggagctgctgacgagatccttgccggcggtgtgaaaggggagcgc 14196
   || || || || || || || || || || || || || || || || || || || || ||
```









D	b		1197	GCGGCGCCNNMCGCCGGCGCGGNCAGCAGCG--CGCCGCGCGCGCCCGC	1140					
Q	y		248	tcatctgttccagacagaccgacacgaacgcgtcgatcgggggtgttcgccgcgcgt	307					
D	b		1139	GCGGCGCGCGCCCCCGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC	1080					
Q	y		308	tctcgccgttcgcgtcgcgcgcgttgtaaacaggfagtgcgcgcgcgcgttgacac	367					
D	b		1079	GCGGC	1020					
Q	y		368	tgc	427					
D	b		1019	C GG GC	960					
Q	y		428	cgcgtcgcgcgcggtcctcgcgctgc--tcgcgtcccgccgcgcgcgcgcgcgcgc	485					
D	b		959	GCGGC	900					
Q	y		486	gcgcgtgtcgc	545					
D	b		899	GCGGC	843					
Q	y		546	ccgc	605					
D	b		842	GC	783					
Q	y		606	gcgcgcgttagcgcgcgagacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	665					
D	b		782	GC	723					
Q	y		666	gatcagtcgcacaagatcgtcaagaogvtgltcacgtlcttccacgcgcacagac	725					
D	b		722	C GCGCGCCNCCC CGC	663					
Q	y		726	gtgtgcacgcacagttgtggccaagcgcgtgtcgtactgtctgtgcgaaccgcga	785					
D	b		662	GNNCGCCCCCGC	603					
Q	y		786	ctctgtccgc	845					
D	b		602	CNNCGC-----GC	548					
Q	y		846	cccgctcgtgtgtaftccaacacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	905					
D	b		547	GCNCGC	488					
Q	y		906	gc-tgc	964					
D	b		487	CCGC	428					
Q	y		965	ggtacgc						
D	b		427	AACACGNGTGG 417						
<b>RESULT 3</b>										
AG032979/c										
LOCUS	AG032979 1452 bp DNA linear GSS 01-NOV-2001									
DEFINITION	Pan troglodytes DNM, clone: PTB-007004.F., genomic survey sequence.									
ACCESSION	AG032979									
VERSION	AG032979.1 GI:16559852									
KEYWORDS	GSS; GSS (genome survey sequence).									
SOURCE	Pan troglodytes male lymphoblast DNM, clone_lib:PTB Chimpanzee Male									
ORGANISM										
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;										
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.										
REFERENCE										
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,										
TITLE Tottohi,Y., Watanabe,H. and Sakaki,Y.										
JOURNAL BAC end sequences of Library PTB										
REFERENCE Unpublished										
2 (bases 1 to 1452)										

AUTHORS	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Shinho-chou, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: shimpbes@sc.riken.go.jp; URL: http://hnp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT	Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors. PRIMERS Sequencing: -21M13 LIBRARY Vector : pKS145 R.Site 1 : SacI R.Site 2 : SacI. Location/Qualifiers 1. 1452 /organism="Pan troglodytes" /db_xref="taxon:9598" /clone="PTB-007004.F" /sex="male" /cell_type="lymphoblast" /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT	23 a 581 c 600 g 74 t 174 others
ORIGIN	
Query Match	8.5%; Score 103; DB 12; Length 1452;
Best Local Similarity	47.0%; Pred. No. 0.00084;
Matches 338; Conservative	0; Mismatches 373; Indels 8; Gaps 2;
Db 142	gacgtgcgcgcggctcctcgcgacgagcgttcgccgacccgacatccgcggttcttg 201
Db 1227	GCGCGNGCGCGCGCNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1168
QY 202	accgacactgtttggcccgcgggtctgtctgcggaatacfccgagcaatatctcttcag 261
Db 1167	GMCGGCGCGCGCGCGC---GGCGCGGMC GGCGCGCGCGCGCGCGCGCGCGCGCG 1111
QY 262	gacgagcccgacaacgacgcgcgcgtgcgcggggtgtgtctgcgcgcgcgcgcgcgc 321
Db 1110	GCNCG 1051
QY 322	ctgcgcgcgcgttgaaacggttgtatcgccgcgcacagtgtgacgcactgtctgcgcgcctg 381
Db 1050	GGCG 991
QY 382	ggccgcgggcgcgatgtgacctgtgttcgacgagctgcgcgtgcataccgccttgcgcgcgcgc 441
Db 990	GGCG 931
QY 442	ctcgcgccttctgcgcctgc 501
Db 930	C CG 871
QY 502	gtggagacgcacccctgacccgcgggcgcgcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 561
Db 870	GGCG 811
QY 562	atccgcgcgagttccgcgcacctaaggtagagcgggcgcgccttcgcgcgcgcgcgcgcgcgcgc 621
Db 810	C CG 751
QY 622	gagagaccttgcgcgttatgtctgcacgcaccaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 681
Db 750	GG 691
QY 682	atcgtcacgacgagtgtaactcatcttcaacgcgcacagagaagcgtgtgcacgcaggtg 741
Db 690	C CGCGCGCGCGCGCG---CCCGCGGNCGCCGCGCMCNMGCCGNCCCGCGCGCGCGCGCGCGCG 636

OY	742	ggcaacgcgcygtcagctctgtcgccaccgcgaacacagcttcgacctctccgagccgg	801
Db	635	ggccggcgccggcccnnnncgccggccccggccggccggccggccggccggccggccgg	576
OY	802	cggagccgctgctggccaggccgctggaggatgctgctgtagtaccgcttgctgagtc	860
Db	575	ggccgcnncggcgccggccgcnccggccggccggccggccggccggccggccggccgg	517
RESULT	4		
LOCUS	AG032943/c		
DEFINITION	Pan troglodytes DNA, clone: PTB-007N10.F, genomic survey sequence.		
ACCESSION	AG032943		
VERSION	AG032943.1 GI:16559816		
KEYWORDS	GSS: GSS (genome survey sequence).		
SOURCE	Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male		
ORGANISM	Pan troglodytes		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.		
TITLE	Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,		
JOURNAL	1 (sites)		
REFERENCE	BAC end sequences of library PTB		
AUTHORS	Unpublished		
TITLE	2 (bases 1 to 1542)		
JOURNAL	Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,		
TITLE	Tokoki,Y., Matanabe,H. and Sakaki,Y.		
JOURNAL	Direct Submission		
TITLE	Submitted (02-AUG-2001) Aseo Fujiyama, The Institute of Physical		
JOURNAL	and Chemical Research (RIKEN), Genomic Sciences Center (GSC);		
TITLE	1-7-22 Suehiro-chou,Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan		
JOURNAL	(E-mail:chlimbes@sc.riken.go.jp, URL:http://bgp.gsc.riken.go.jp/,		
TITLE	Tel:+81-45-503-9111, Fax:+81-45-503-9170)		
JOURNAL	Clones are derived from the chimpanzee BAC library PTB This BAC end		
TITLE	was generated during the R&d process and may have higher chance of		
JOURNAL	clone tracking errors.		
COMMENT	PRIMERS		
TITLE	Sequencing: -21M13		
JOURNAL	LIBRARY		
TITLE	Vector : pKS145		
JOURNAL	R.Site 1 : SacI		
TITLE	R.Site 2 : SacI.		
JOURNAL	Location/Qualifiers		
TITLE	1..1542		
JOURNAL	/organism="Pan troglodytes"		
TITLE	/db_xref="taxon.9598"		
JOURNAL	/clone="PTB-007N10.F"		
TITLE	/sex="male"		
JOURNAL	/cell_type="lymphoblast"		
TITLE	/clone_lib="PTB Chimpanzee Male BAC Library"		
JOURNAL	BASE COUNT		
TITLE	ORIGIN		
JOURNAL	61 a 657 c 614 g 19 t 191 others		
Query Match	8.4%; Score 101.8; DB 12; Length 1542;		
Best Local Similarity	45.9%; Pred. No. 0.0012;		
Matches 512; Conservative	0; Mismatches 590; Indels 14; Gaps 8;		
OY	27	cgatcccaacgacgcgcgaagctgagcgctgacccctaaccgctcttaacactgctgctcg	86
Db	1445	CGCGCCGGGCGGCCGCCTGGCCGCGCAGCCGCGCCGCGCCGCGCCGCGCCGCGCCG	1386
OY	87	gcacgaaccgctgacccgctgagcccaaccggtctcgtgacgtctcccgcttcggagaagt	146
Db	1385	GGCGGCGNCNGCG	1326
OY	147	gagcgcgctgctgcgcgcgaagcgcttcgcgcgcgaacgcgatccgcgcgcttcggaacga	206
Db	1325	GGCG	1267
OY	207	ctctgctggccgcgggctgctctgcgcagatagctcgcgacatatcatctctttccagaa	266

[illegible]



ORGANISM Magnaporthe grisea  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Sordariomycetes Incertae sedis; Magnaportheaceae; Magnaporthe.

REFERENCE 1 (bases 1 to 1516)  
AUTHORS Choi, W. and Dean, R.A.  
TITLE Construction and sequence analysis of an appressorium stage cDNA library in the rice blast fungus, Magnaporthe grisea  
JOURNAL Unpublished (2001)  
COMMENT Contact: Ralph A. Dean  
Fungal Genomics Laboratory  
North Carolina State University  
Campus Box 7251, Raleigh, NC 27695, USA  
Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph.dean@ncsu.edu  
Seq primer: T3 primer (AATTAACTCTACTTAAGG).

FEATURES  
source  
1..1516  
/organism="Magnaporthe grisea"  
/strain="70-15"  
/db\_xref="taxon:148305"  
/clone="mgct002xdl1f"  
/clone.lib="Magnaporthe grisea Appressorium Stage cDNA"  
/dev\_stage="Germinated conidia on appressorium-inductive surface"  
/note="Vector: pBluescript SK(+) Vector; Site\_1: EcoRI; Site\_2: XhoI; The appressorium formation-specific cDNA library was constructed from conidia germinated for 5-8 hr on an inductive surface. The library contains over 55,000 clones with average insert size of 1.5 kbp."

BASE COUNT 65 a 657 c 660 g 16 t 118 others  
ORIGIN

Query Match 8.4%; Score 101; DB 10; Length 1516;  
Best local Similarity 43.9%; Pred. No. 0.0014;  
Matches 408; Conservative 0; Mismatches 516; Indels 5; Gaps 2;

QY 39 cgcgcagctgacgagcgtgacccctaccctgctaccactgctctgacgacaccggt 98  
DB 1107 CGCGCMCGGGGGGGCGCCNMCNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1048  
QY 99 gacgcgtgagccacacgggtctgtgactctccgcttcgcgacgtgacgagtgct 158  
DB 1047 GCGCGGGGGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 988  
QY 159 cgcgcagacgagcgttcgcccgcgacgacatccgcggtcttgacacgactgctg 218  
DB 987 CG 928  
QY 219 cgggctgctgcgcgagatcgtcgcgacacatcctcttcacgagacgacccagcg 278  
DB 927 GCG 868  
QY 279 ccgagtcgcgcgaggtgtgtcgcgcgcggttcttcgcgcttcgcgctgacgac 338  
DB 867 GCG 808  
QY 339 ggtgacgcgcgacgacgctgtagacacgctgtgcgcccgcgcgcgcgcgcgacatga 398  
DB 807 CG 748  
QY 399 cgtgtgtagacgagctgtagcgttacgcgctgagcgtgcggtcttcgacctgacct 458  
DB 747 CCCGNNCGCGNCG 688  
QY 459 gccgcgcgcgacgagtggtgagcgtgctgcgacgacgtgtagacgacacctga 518  
DB 687 GCG 631  
QY 519 ccgagtcgcgcgagcagacatgtagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 578  
DB 630 CNGGCG 571

QY 579 ctacgtgagagcgcccttcgcgagcgccgctgaggtcgagacactgtgacct 638  
DB 570 GCG 511  
QY 639 gatctcgagcgcacacgacccgcgctgtagtcgaacgagatcgcacagagtggt 698  
DB 510 CG 451  
QY 699 cagcttacccttcacgc 756  
DB 450 CCG 391  
QY 757 agcctgctgagcgcacgc 816  
DB 390 CCG 331  
QY 817 cagcgcgtcgagagtgctgcgtgtagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 876  
DB 330 GCG 271  
QY 877 gacgtcgagctgagctgc 936  
DB 270 GCG 211  
QY 937 gccgc 965  
DB 210 GCG 182

RESULT 6  
BG809984 1516 bp mRNA Linear EST 22-MAY-2001  
LOCUS  
DEFINITION mgct002xdl1f Magnaporthe grisea Appressorium Stage cDNA Magnaporthe  
grisea cDNA clone mgct002xdl1f 5', mRNA sequence.  
ACCESSION BG809984  
VERSION BG809984.1 GI:14180964  
KEYWORDS EST.  
SOURCE  
ORGANISM Magnaporthe grisea.  
Magnaporthe grisea.  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Sordariomycetes Incertae sedis; Magnaportheaceae; Magnaporthe.  
REFERENCE 1 (bases 1 to 1516)  
AUTHORS Choi, W. and Dean, R.A.  
TITLE Construction and sequence analysis of an appressorium stage cDNA library in the rice blast fungus, Magnaporthe grisea  
JOURNAL Unpublished (2001)  
COMMENT Contact: Ralph A. Dean  
Fungal Genomics Laboratory  
North Carolina State University  
Campus Box 7251, Raleigh, NC 27695, USA  
Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph.dean@ncsu.edu  
Seq primer: T3 primer (AATTAACTCTACTTAAGG).

FEATURES  
source  
1..1516  
/organism="Magnaporthe grisea"  
/strain="70-15"  
/db\_xref="taxon:148305"  
/clone="mgct002xdl1f"  
/clone.lib="Magnaporthe grisea Appressorium Stage cDNA"  
/dev\_stage="Germinated conidia on appressorium-inductive surface"  
/note="Vector: pBluescript SK(+) Vector; Site\_1: EcoRI; Site\_2: XhoI; The appressorium formation-specific cDNA library was constructed from conidia germinated for 5-8 hr on an inductive surface. The library contains over 55,000 clones with average insert size of 1.5 kbp."

BASE COUNT 65 a 657 c 660 g 16 t 118 others  
ORIGIN

[illegible][illegible]

[illegible][illegible]









[illegible][illegible]





.

OM of: US-09-724-797-36 to: GenEmbl:\* out\_format : pfs  
Date: Jun 21, 2002 6:12 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:  
-MODEL-frame+ p2n.model -DEV=xlh  
-O/cgml2.1/USPTO.spool/US09724797/runat.21062002.092539.18314/app-query.fasta.1.466  
-DB-GenEmbl -OFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.500  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500  
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-DELEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blissum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZ=500  
-MINLEN=0 -MAXLEN=200000000 -USER=US09724797\_@CGML\_1\_9912  
-NCPU=6 -ICPU=3 -LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30  
-NO\_XLPHY -WAIT -THREADS=1

## Search information block:

Query: US-09-724-797-36

Database: GenEmbl:\*

Database sequences: 1797656

Database length: 187333701

Search time (sec): 1850.910000

Sequence	Strd	Orig	ZScore	Escore	Len	Documentation
gb_pat:AR173058	+	609.00	595.36	8.4e-25	8478	AR173058 Sequence 1 from patent
gb_da:BSU51868	-	609.00	593.86	1.0e-24	10153	U51868 Bacillus subtilis bio
gb_da:BSU60016	-	609.00	568.34	2.7e-23	215640	I 299119 Bacillus subtilis com
gb_da:AF008220	+	609.00	568.17	2.7e-23	220060	I AF008220 Bacillus subtilis r
gb_da:SC10	-	572.50	546.25	4.5e-22	39524	AL049754 Streptomyces coelic
gb_da:AB070947	-	563.00	540.88	9.0e-22	24268	AB070947 Streptomyces avermit
gb_da:AE004675	-	561.50	546.75	4.3e-22	10060	I AE004675 Pseudomonas aerugin
gb_da:AF087022	+	538.50	539.96	1.0e-21	1470	I AF087022 Streptomyces venezuel
gb_da:SEPC9450A	+	538.50	530.91	3.2e-21	4342	I AF079139 Streptomyces venezuel
gb_da:MLERTN8	-	531.00	526.34	3.8e-21	3078	I M63110 Saccharopolyspora eryth
gb_da:CE34	-	520.00	477.06	3.2e-18	342300	I AL583924 Mycobacterium lepra
gb_da:AE046080	-	520.00	494.87	3.3e-19	36028	I AL442120 Streptomyces coelic
gb_da:AE0304	+	509.50	470.55	7.5e-18	190050	I AL646080 Ralstonia solanace
gb_pat:AE0304	+	508.50	498.29	2.1e-19	6085	I A60394 Sequence 4 from Patent
gb_pat:AR144763	+	508.50	478.68	2.1e-19	6085	I AR144763 Sequence 4 from Patent
gb_da:AF127374	+	501.00	479.92	1.4e-18	63734	I AF127374 Streptomyces lavendu
gb_da:AF254925	+	501.00	501.53	1.4e-19	12511	I AF254925 Streptomyces maritim
gb_pat:AX094757	+	498.50	480.96	2.0e-18	14759	I AX094757 Sequence 22 from Pat
gb_da:AE004755	+	490.50	482.04	1.7e-18	5008	I AE004755 Pseudomonas aerugin
gb_da:BSU93876	-	490.50	471.43	6.7e-18	17828	I U93876 Bacillus subtilis amir
gb_da:BSU80014	-	480.50	464.26	1.5e-17	213420	I Z99117 Bacillus subtilis com
gb_da:AE003889	-	478.50	464.26	1.5e-17	10092	I AE003889 Xylella fastidiosa g
gb_da:SEKRTFGH	+	477.00	475.33	4.1e-18	2243	I M54983 Saccharopolyspora eryth
gb_da:SEKRTFBI0	+	472.50	466.01	1.3e-17	4254	I X60379 S. erythraea erythromyci
gb_da:STMOLEP	+	472.50	474.80	4.3e-18	1400	I L37200 Streptomyces antibiotici
gb_pat:AR159871	+	471.50	444.79	2.0e-16	50937	I AR159871 Sequence 1 from Pat
gb_da:MPUNYCG	+	471.50	470.15	7.9e-18	2168	I D16098 Micromonospora griseoru
gb_da:E08092	+	471.50	470.15	7.9e-18	2168	I E08092 Genomic DNA encoding my
gb_da:BSU80010	-	471.00	430.57	1.3e-15	233780	I Z99113 Bacillus subtilis com
gb_da:AE003887	+	467.50	453.80	6.9e-17	10261	I AE003887 Xylella fastidiosa g
gb_da:STMSUBCB	+	463.00	437.63	5.2e-16	39228	I M32239 S. griseolus cytochrome
gb_da:MLC81788	-	463.00	437.63	5.2e-16	39228	I AL06869 Mycobacterium leprae
gb_da:SHGCPFR	-	461.50	427.63	1.8e-15	107329	I X60780 S. hygroscopicus gene
gb_da:AB018074	+	461.00	457.60	3.9e-17	2795	I AB018074 Streptomyces griseus
gb_pat:ES9713	+	459.50	432.69	3.9e-17	2795	I E59713 Process for microorgan
gb_da:SC8811	-	459.50	449.55	1.1e-16	5774	I AL53384 Streptomyces coelic
gb_da:AB070948	+	457.50	456.21	4.7e-17	2177	I AB070948 Streptomyces avermit
gb_da:AF102510	-	456.00	437.67	5.1e-16	16775	I AF102510 Mycobacterium smegmat
gb_da:AB000735	-	456.00	437.67	5.1e-16	16775	I AB000735 Nocardioides sp. gen

gb_da:AE007076	-	446.00	428.56	1.6e-15	15201	I AE007076 Mycobacterium tub
gb_da:HTC339	-	446.00	419.90	5.0e-15	42861	I Z77163 Mycobacterium tuber
gb_da:FE015825	+	444.50	409.93	4.9e-15	35739	I AF015825 Bacillus subtilis
gb_da:BSU80007	+	444.50	404.88	3.4e-14	216750	I Z99110 Bacillus subtilis
gb_da:AB036790	-	444.00	437.33	5.3e-16	4192	I AB036790 Microcystis aerugi

seq\_name: gb\_pat:AR173058

## seq\_documentation\_block:

LOCUS	AR173058	8478 bp	DNA	linear	PAT 17-DEC-2001
DEFINITION	Sequence 1 from patent US 6303377.				
ACCESSION	AR173058				
VERSION	AR173058.1	GI:17912549			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				

## REFERENCE

1 (bases 1 to 8478)

Bower, S. Grant, Perkins, J. B., Yocum, R. Rogers, and Petro, J. G.

Biotin biosynthesis in Bacillus subtilis

Patent: US 6303377-A 1 16-OCT-2001;

Location/Qualifiers

1..8478

/organism="unknown"

BASE COUNT 2432 a 1749 c 2088 g 2204 t 5 others

ORIGIN

## alignment\_scores:

Quality:	609.00	Length:	396
Ratio:	2.316	Gaps:	5
Percent Similarity:	66.414	Percent Identity:	33.586

## alignment\_block:

US-09-724-797-36 x AR173058 ..

Align seg 1/1 to: AR173058 from: 1 to: 8478

8	AlpheasProthraspalaaspvalAArgaaspProtyrProSerty	24
5493	GCATCGCAACATCACTCTGAGTTTGGAAAACCCATATCTTTTAA	5542
24	rhISTPLeuAargHIsasproValHIsargVala.....H	38
5543	CGACATTCGCGAGCTTCATCTATTAAGGAGCTTCTTAAT	5592
38	isArgValITryrValserArgpheAlaaspValaArgAlaValleugly	54
5593	ACCGGCGCTGATGCACAGATGAGAAAGCGCTCTATTGGAA	5642
55	AspGluArgphe...AlaArgHrglyLeaArgaArgpHrTrpThasple	70
5643	GATGCGAGATCAAGATCCGACCCGCTGAGAGCTCAACCAATA	5692
70	uValGlyProGlyLeuLeuAlaGluLeuValaGlyaspIleleuPheg	87
5693	TCAGGAC.....CTTCACATGTGCAAAATCAATGAGTCTGTTT	5733
87	lnAspGluProAspHisGlyArgLeuArgGlyValaGlyProAlaPhe	103
5734	AGAACAGGCTGATCATGACGATTCGCGAGCTGCGAGCGAGCTTT	5783
104	SerProSerAlaLeuAargArgLeuGluProValIleAlaGlyThValas	120
5784	ACGCCGAGAACACAGAGCTTATCAAGCCGATTCATTAAGAACTGCA	5833
120	pAspLeuAargProAlaLeuAlaArgGlyValaMetaspValaIaPsp	137
5834	TCATTGCTTGTATCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA	5883
137	lnLeuAlaArgProLeuAlaLeuAlaValaLeuGlyLeuLeuGlyLeu	153
5884	ACTTTCCTTTCTTACCAAGTTTGTCAATGATTAAGTATAGTGCTA	5933

154 ProAlaAlaAspTrpGlyAlaValGlyArgTrpSerArgAspValGlyArg 170  
170 gTrrLeuAspArgGlyAlaSerAlaGluAspMetArgGlyAlaAla 187  
5984 AACGATGATGTTTACCCTCAGAAAGGATTAACAGAGGCAATATTA 6033  
187 laIleAlaGluPheAlaAspTyrValGluArgAlaLeuAlaArgArg 203  
6034 TGCGTGTGACAGGCTATGGCATATTTCAAGAGCTGATTTCAAAAGAGAAA 6083  
204 ArgGluGlyGlyLeuAspLeuAlaLeuMetLeuAspAlaHisAsp 220  
6084 CGCCACCCCTACAGAGATATGATCAGCATCTCTTGAAGGGGAGAGAAAA 6133  
220 gGlyLeuMetSerArgAsnGluIleValSerThrValAlaThrPheIleP 237  
6134 GGATTAACCTGACGAAAGAGAGGCGGCATCTACGTGATATTCGTGGCA 6183  
237 heThrGlyHisGluThrValAlaSerGluValGlyAsnAlaValLeuSer 253  
6184 TCCGCGGACATGAGACACACGTCATCTCATTCAGCAATTCAGTCCTTGT 6233  
254 LeuLeuAlaHisProAspGluLeuAspLeuLeuArgArgArgProAsp 270  
6234 CTGCTGACGATCCAGAACAGCTTTTGAACCTGAGAGAAATCAGATCT 6283  
270 uLeuAlaGluAlaValGluGlyLeuArgTyrAspProSerValGlns 287  
6284 TATTGTACCGCAGCTCGAGAAATGTTTACGCTATGAAGCCCAACGCAAA 6333  
287 eraSntTrpArgGluLeuAspValAlaGluLeuArgGlyArgArgLeu 303  
6334 TGACAGCAGAGTTGCGTCAGAGATATGACATCTCGGGGTGACGATC 6383  
304 ArgArgAspValValAlaValAlaLeuAlaGluAlaAsnArgAsp 320  
6384 CGTACAGAGAGAACAGTCTATCTTTTGTAGAGCGGCTATGACAGACC 6433  
320 oArGArgTyrAspArgProAspAspPheAspIleGluArgAspProValP 337  
6434 TAGCATATTCAGAACCCCGATGCTTCGATATACGAGAAAGTCTATATC 6483  
337 roSerMetSerPheGlyAlaGlyMetArgTyrCysLeuGlySerTyrLeu 353  
6484 CGCATCTTTTCATTCGGGATGGCCATCATGTTTCTTAGGGTCTCGCTG 6533  
354 AlaArgThrGluLeuArgAlaAlaVal...AlaAlaLeuAlaArgLeuP 369  
6534 GCAGATTAAGAGCGCAATGCGATTAACACTCTTCGACGCAATGCC 6583  
369 oGlyLeuArgLeuGlyCysAlaSerAspAlaLeuAlaTyrGlnProArgT 386  
6584 CAGCCTTATCTT...CGCGATTTTGAATGGCGGATCGCGGCTTT 6627  
386 hmePheArgGlyLeuAlaSerLeuProIleAlaPhe 398  
6628 TTGGATTTGCGGGCTTGAGAGACTGCGGTGACTTTT 6665

seq\_name: gb\_ba:BSU51868

seq\_documentation block: 10153 bp DNA linear BCT 25-OCT-1996

LOCUS BSU51868

DEFINITION Bacillus subtilis biotin biosynthetic operon genes, complete and partial cds.

ACCESSION U51868

VERSION U51868.1 GI:1277024

KEYWORDS

SOURCE Bacillus subtilis.

ORGANISM Bacillus subtilis

Bacteria; Firmicutes; Bacillus/Clostridium group;

REFERENCE 1 (bases 1 to 10000)  
AUTHORS Bower, S., Perkins, J.B., Yocum, R.R., Howitt, C.L., Rahaim, P. and Pero, J.  
TITLE Cloning, sequencing, and characterization of the Bacillus subtilis biotin biosynthetic operon  
JOURNAL J. Bacteriol. 178 (14), 4122-4130 (1996)  
MEDLINE 96312354  
REFERENCE 2 (bases 1 to 10153)  
AUTHORS Yocum, R.  
TITLE Direct Submission  
JOURNAL Submitted (20-FEB-1996) Yocum R., Omigene Bioproducts, Inc., 763 D Concord Ave., Cambridge, MA, USA, 02139-9002  
FEATURES  
source location/Qualifiers  
1..10153  
/organism="Bacillus subtilis"  
/db\_xref="taxon:1423"  
/map="262 degrees"  
/clone="PB10100, PB10350, PB10201"  
2119..2898  
/gene="biow"  
2119..2898  
/gene="biow"  
/function="biotin pathway"  
/codon\_start=1  
/transl\_table=11  
/evidence="experimental"  
/product="pimelyl CoA synthetase"  
/protein\_id="AAB17457.1"  
/db\_xref="GI:1277025"  
/translation="MQEETFYVRKASNSGHEDGKHISGGERLIPFEHKHTVN ALLEKGSISHRGKPRDEMQIOFEVHESIKTIQPLVPHNVEVSCPEGOKLARILLEKE GVSRDVLEKAYEQIPEMSDVGAVALFDIHKRMDOETKEGVASRMDPADFEKWA LHSVPAHSRIKALALASKVSRRHPAVVAELCWSDDPDYITGVAKKMGYORITAMK EYGECECRVFEFDIGSNDVNTYIHDEKQPIILEWEDHDS"  
2888..4234  
/gene="biow"  
2888..4234  
/gene="biow"  
/function="biotin pathway"  
/codon\_start=1  
/transl\_table=11  
/evidence="experimental"  
/product="DAPA aminotransferase"  
/protein\_id="AAB17458.1"  
/db\_xref="GI:1277026"  
/translation="MTHDLIEKSKHMLPFTQMKDYDENPLIIESGTGIRKVDINK EYDGFSSVWLNVHGHRKELEDAIKKQIKIAHSTLLGWTNPATQLETLIDISPK KLPVFEVSDGAEFAMEITALKMARQYWNKIKPEKOFIAMNGYHGDTICAVSGSIE LFNHYVPIPMPESEKAPITPVYVRESGDPCECDQCLRELAOLLEHHEHRIALSTES MVQASGMIVMPEBYLAGVBELCTTYVLMITVBVATGRTGKMPACRENNOPDLM AAGGIGTGYLPIAVTFATEDIYKAFYDDYENLTFPHGVSYTGNOIGCAVALENLAL FESENIEVOYVAEKSKLHFLDIDLHALPHVGDIRQLDFMCGAELVNSKREYEPADR RIGKYSKMKMELMLTRPLGDVIAFLPPLASTAEELSEVYALMKAHIEVTSLED"  
4224..5393  
/gene="biow"  
4224..5393  
/gene="biow"  
/standard\_name="KAPA synthase"  
/function="biotin pathway"  
/codon\_start=1  
/transl\_table=11  
/evidence="experimental"  
/product="L-alanine - pimelyl CoA ligase"  
/protein\_id="AAB17459.1"  
/db\_xref="GI:1277027"  
/translation="MKTDSWNERLDHMKKEAGVHRLRSMGAPVPERINDGENQTVW SSNNITGLASRRILDAQRTALQOPFGSSGRSLTNGNSVHEKLEKXIASPKLLEA LIFSSTGLANGLVLSLPEKEEDVILSQLNHASKITDCRSLKADTVYVYRIIDNDLEN KLNFORQORRFIVTDGVFSMDGTIAPLDQIISLAKYHAFVAVDDNHAATGADSDQ GTSEFVGPDIIVIGTLSKAGAGGAGAGSAVFIDPLNHAFTFIQTALIPASCA AHEAFNITEASREKRLDFSYISMRISLKMRYGVKGDHPIIPVYIGAHNTVFA EKLOGKGIYAIRPPIYVAPESRIRITITISDSHMDIDLQTFHSGKELHI1"

```
gene
5390..6085
/ gene="biod"
/ product="unknown"
/ protein_id="AAB17461.1"
/ db_xref="GI:1277032"
/ translation="MGAT1AHSKLNRLFDMLYGFLLMALICVLPFTHVIAASFATVEVSKFKLIPPTFESIDARFYFESDIIKSLVSEFVTVGTAVSMFLSLMAGLSRDLIGROPIMFLVFTMFSSGMIPTFLVYKSLDLSVWALILPTAINFNLILKNEFQNPSSLEBSAKIDGNDIGIFKFLVLPISLPAIATISLFAVTVMTYMTAILYLMSAKMPIOVLLROIYIVSSGMDMSGSPPEQTIG"
/ base_count 2886 a 2093 c 2502 g 2672 t
ORIGIN

alignment_scores:
  quality: 609.00      length: 396
  ratio: 2.316         gaps: 5
  percent similarity: 66.414  percent identity: 33.586

alignment_block:
US-09-724-797-36 x BSU51868 ..

Align seg 1/1 to: BSU51868 from: 1 to: 10153

8 AlphaheaprophrAspAlaaspValArgArgaspprotyrProserly 24
||| ..:.....:|||||
7173 GCATCGCACTGCATCTCTGAGTTTGGAAAACCATATTCCTTTTA 7222
24 rHs1rpleuLeuArgHisaspproValHisArgGlyAla.....H 38
||| ||| ..:.....:|||||
7223 CGACACATTCGAGACCTGTCATCCTATCTATAAGGAGTTCTTAAAT 7272
38 IsArgVal1rPtyrValSerArgPheAlaaspValArgAlaValLeuGly 54
::: ||||| ||||| ..: ..: ..: ||||| |||||
7273 ACCCGGCTGTATGTCACAGATATGMAAGAAAGCGCTGATATTGAAA 7322
55 AspGluArgPhe...AlaArgThrGlyLeuArgArgPheThrAspLe 70
||| ||||| ..: ||||| ..: ..: |||||
7323 GATGCGAGATTCAAGTCCGACCCCGCTGCCGAGACGTCACCAATAA 7372
70 vAlaGlyProGlyLeuLeuAlaGluIleValGlyAspIleIleLeuPhe 87
||| ||||| ..: ||||| ..: ..: |||||
7373 TCAAGAC.....CTTTCACATGTGCMAATTAAGATGCTGTTTC 7413
87 lnsAspGluProAspHisGlyArgLeuArgGlyValValGlyProAlaPhe 103
||| ||||| ||||| ||||| ..: ..: |||||
7414 AGAACCGGCTGATCAATAGCATGCGAGCGCTGCCAGCGAGCGTTT 7463
104 SerProSerAlaLeuArgArgLeuGluProValIleAlaGlyThrValAs 120
::: ||| ..: ..: ||||| |||||
7464 ACCCGGAGAACGACAGAGATTATCAGCCGTATATCATTTGAACAGTCCA 7513
120 pAspIleuLeuArgProAlaLeuAlaArgGlyAlaMetAspValValAsp 137
||||| ||||| ..: ..: ||||| |||||
7514 TCATTTGCTTGATCAAGTGCMAAGTGAATAAAGAGTGGGTCATTTCCG 7563
137 lueuAlaTyrrProLeuAlaLeuArgAlaValLeuGlyLeuGluGlyLeu 153
::: ||||| ||||| ..: ..: ||||| |||||
7564 ACTTGCTTTTCTCTTAGCAGATTGTGTCATAGCTACATTTAGTGTA 7613
154 ProAlaAlaAspTrrpAlaValAlaGlyArgTrrpSerArgAspValGlyAr 170
||| ||||| ..: ..: ||||| |||||
7614 CCGAGAGAAAGATAGGAGCAATTAAAGAGTGGGTCGAGATCTCATCCA 7663
170 gThrLeuAspArgGlyAlaSerAlaGluLysPheIleArgGlyHisAlaAla 187
||| ||||| ..: ..: ||||| |||||
7664 AACGATTGATTTTACCCTGCAAGAAAGCATTAACAGAGGCAATATTA 7713
187 lAlleAlaGluPheAlaAspTyrrValGluArgAlaLeuAlaArgArgArg 203
::: ||| ..: ..: ||||| |||||
7714 TGGCTGTGAGGCTATAGCATATTTCAAGAGCTGATTCMAAAGGAAAA 7763
204 ArgGluGlyGlyLusPheLeuAlaLeuMetLeuAspAlaHisAspArg 220
||| ||||| ..: ||||| ..: ..: |||||
```

```

7764 CGCCACCCCTCAGACAGATATGATTCAGCATGCTCTTGAAGGGAGAGAAAA 7813
220 ggilyleuMetSerArgAsnGluIleValSerThrValValThrPheIleP 237
7814 GGATTAAGCTCAGCAAGAGAGCGGCGATCTACGCGCATATGCTTGCGCA 7863
237 heThrglyHisGluThrValAlaSerGlnValGlyAsnAlaValIleSer 253
7864 TCGGCGGACATGAGACACAGCTCAATCTCATCAGCAATTCAGTCTTGT 7913
254 leuLeuAlaHisProAspGlnLeuAspIleuLeuArgArgProAspIle 270
7914 CTGCTGCAGCATCAGACACAGCTTTGAAACAGAGAAAAATCCAGATCT 7963
270 uLeuAlaGlnAlaValGluGluCysLeuArgTyrAspProSerValGlns 287
7964 TATTGGTACCGCAGCTGAGGAATGTTTACGCTATGAAGCCCGACGMA 8013
287 eArSnThrArgGlnLeuAspValAspValGluLeuArgGlyArgArgLeu 303
8014 TGACAGCCAGAGTTCGCTCAGAGATATTGACATCTCGCGGGTACAGATC 8063
304 ArgArgAspAspValValValValIleuAlaGlyAlaAlaAsnArgAspR 320
8064 CGCTAAGGAGACACAGCTATCTTTTGTTAGGAGCGGCTAATCCAGACC 8113
320 cArgArgTyrAspArgProAspAspPheAspIleGluArgAspProValP 337
8114 TAGCATATTCACGACACCCGATGCTTTCATATTCAGAGAGTCCATATC 8163
337 roSerMetSerPheGlyAlaGlyMetArgTyrCysLeuGlySerTyrIleu 353
8164 CGCATCTTTTCATTCGGGCATGGCCATCATGTTGCTTAGGGCCCTCGGTG 8213
354 AlaArgThrGlnLeuArgAlaValAlaVal...AlaAlaLeuAlaArgLeuP 369
8214 GCAGAGATTAGACGCAAAATTCGATTAACACTCTTCGACGCAATGCC 8263
369 cGlyLeuArgLeuGlyCysAlaSerAspAlaLeuAlaValArgGlnProArgT 386
8264 CAGCCTTAATCTT....CGCGATTTCGAATGGCGGTATCGCGCGCTTT 8307
386 hrMetPheArgGlyLeuAlaSerLeuProIleAlaPhe 398
8308 TTGGATTTCGGCGCCTTGAGAGACTCGCGGTACATTTT 8345
seq_name: gb_ba:BSUB0016

```

self document	Accession	215540 bp	DNA	linear	BCT 26-NOV-1997
LOCUS	BSUB0016				
DEFINITION	<i>Bacillus subtilis</i> complete genome (section 16 of 21) : from 2977771 to 3213410.				
ACCESSION	29919 AL009126				
VERSION	29919.1	GI:2635411			
KEYWORDS					
SOURCE	<i>Bacillus subtilis</i> .				
ORGANISM	<i>Bacillus subtilis</i>				
REFERENCE	<i>Bacteria</i> ; Firmicutes; <i>Bacillus</i> /Clostridium group; <i>Bacillus</i> /Staphylococcus group; <i>Bacillus</i> .				
AUTHORS	1 (bases 1 to 215640) Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bartero, M.G., Bessieres, P., Bolotin, A., Borchert, S., Borriis, R., Boudier, L., Brans, A., Braun, M., Brignell, S.C., Brois, S., Brouillet, S., Brusch, C.V., Caldwell, B., Capuano, V., Carter, N.M., Choi, S.K., Codani, J.J., Conerton, I.F., Cummings, N.J., Daniel, R.A., Denizot, F., Devigne, K.M., Dusterhoft, A., Ehrlich, S.D., Emmerston, P.T., Entian, K.D., Errington, J., Fabre, C., Ferrier, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galloni, N., Chim, S.Y., Glaser, P., Goffeau, A., Golligly, E.J., Grandi, G., Guiseppl, G., Guy, B., J. Haga, K., Hated, J., Harwood, C.R., Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Kraerr-Blanchard, M., Klein, C., Kobayashi, Y., Koeltter, P.,				

TITLE	The complete genome sequence of the gram-positive bacterium <i>Bacillus subtilis</i>
JOURNAL	Nature 390 (6657), 249-256 (1997)
MEDLINE	98044033
REFERENCE	2 (bases 1 to 215640)
AUTHORS	Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
TITLE	Direct Submission
JOURNAL	Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur, Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75722 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr, adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48

source 1. .215640  
 location/Qualifiers  
 /organism="Bacillus subtilis"  
 /strain="168"  
 /db\_xref="taxon:1423"  
 complement(8. .1001)  
 /gene="ytm"  
 complement(8. .24)  
 /gene="ytm"  
 complement(99. .1001)  
 /gene="ytm"  
 /function="unknown"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAB14906.1"  
 /db\_xref="gi:2635412"  
 /translation="MKKLIVFAFISGLIDSGIGMAYGVSTSLLAFGITPAVAS  
 ASVLAEEVTTAAAGSVSHIKFGNDKQVYOLVIGSIGAFGLAAFLSQLDGDAKPP  
 ISLELLIGGVYLIRFLFOYKPALEKKHVPVLRKRSIPLGIYAGADATGGGGGKPV  
 TPILSRKGLSRKRYGVTVDSFPAIANSATAGSLISIGMEVDNMLWFSLMAGIITR  
 APIAMVLQKHPQDLMGVLVGGFIIIVNARILINEMIANTVNHPILITAIATMLSNV  
 LEVLSIKSNRNVITKAVDVHLK"  
 complement(1020. .2270)  
 /gene="hipo"  
 complement(1020. .2270)  
 /gene="hipo"  
 /EC\_number="3.5.1.32"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hipurate hydrolase"  
 /protein\_id="CAB14907.1"  
 /db\_xref="gi:2635413"  
 /db\_xref="SPTREMBL:O34980"  
 /translation="MSIDYWRNIIEGSSYPYQTTGNDILFLKEESPVNLTLEKOLIG  
 RHHQIYPLSKEEFTTAFRIKCLKKNGIQRPAALTKGVPAIDAGSSEPAIALRR  
 DIDDALIEKKTGLPYASKKHGIMHACGHDFHTAALGAAFLIKENODSLKRIETB  
 PAEEAGAGATYKIEDGQLDGIDAVIGLHNKPDIAVIGTGPLAKAYDNRKVIETB  
 KAHVALPENGFDPIIGASOLIVALQITVSRVNPLOSAILTVGKINGSTWNPVLPD  
 VYIESTVRFESEVRNQVKQREFAVTEQISAFSLKANWKMSGPPCLCNDEATIGLL  
 RQAARHAKLOVIDPAPSTAGEDFAFYLEHIIPGSEAFEGTGDGDHDMHNPATFIDETAII  
 KSYLTVESAKRLIDSNESKISD"  
 complement(2288. .2980)  
 /gene="tibr"  
 /gene="tibr"

```
CDS      complement(2288. .2980)
          /gene="ribB"
          /function="regulation of riboflavin biosynthesis genes"
          /note="alternate gene name: ytnK"
          /codon_start=1
          /transl_table=11
          /product="transcriptional regulator"
          /protein_id="CAB14908.1"
          /db_xref="GI:2635414"
          /db_xref="SWISS-PROT:P94465"
          /translation="MTIAGTVKKGKQGLKGFPTANYDAKIHGLRNGVYGLATVN
          HQFHLAGVMNIGVAPVGSNLEKTLLEIFEDFHDYGEKIECSILFKIIEERPSLE
          SLTKQIKDKDISCAKPEFLIGINAPRKESLSHQLNPDLCFYKKCNMLGVKRGV
          YNVIDMPEFVGTVAVRIRYILSFLSKEDNPKVSSKIRFGAGGLADLNKPFIS
          SYVESEENLIG"
          complement(3027. .4355)
          /gene="ytnJ"
          complement(3027. .4355)
          /function="unknown"
          /note="similar to nitrlolriacetate monooxygenase"
          /codon_start=1
          /transl_table=11
          /protein_id="CAB14909.1"
          /db_xref="GI:2635415"
          /db_xref="SPTREMBL:O34974"
          /translation="MTRADFIQFAMIHGVGGTTPDGMRRHVDVPSASTNIEPYMKKQ
          TAEKGLSFIFDIADGLFISEKISPHPLINPEPITLISALASTKNIGLVGTSTFTE
          PPTISRLMSLDHISGRAGKMLVTSPOGAKRNKSKMLPHTIRTEIJAQSHLDVVR
          GIMNSWEHDAFTHNKKTGFQFOAKLRLNHKGYEQVEGPLNIGSRKQGEFVIVQ
          SSETGRQFAKANADALEPHNSNLEETKAFYADVKSRADEGDPSSVRIFPDISIVA
          DTEEAERKYYREFAEILIPENAVYTLARFDDYDLISVPLDEFPDIDIGVKNACOST
          DTRIKREAKARNLIREVAQEMAFPEPTLEIGPEVVASILETFMNAFADGFIVSDDI
          PGTLFAFVEKVIPIIOERGLRQDYGRTGLRNLGLGIPQHSVLSHSH"
          complement(4352. .4633)
          /gene="ytnI"
          complement(4352. .4633)
          /function="unknown"
          /codon_start=1
          /transl_table=11
          /protein_id="CAB14910.1"
          /db_xref="GI:2635416"
          /db_xref="SPTREMBL:O34639"
          /translation="MSDVNIYVMSKKGSCYCEEVKNYLNENKFPFONIDYSEKELR
          DILQVYVRAHPVVEIGRNOYOGITELGIEHLALANHAQIKEAR"
          complement(4648. .5652)
          /gene="ytnO"
          complement(4648. .5652)
          /function="unknown"
          /note="similar to hypothetical proteins"
          /codon_start=1
          /transl_table=11
          /protein_id="CAB14911.1"
          /db_xref="GI:2635417"
          /db_xref="SPTREMBL:O34846"
          /translation="MIRLSILDPSLIGGETAOTLOHTVVKLAQMAEEGCGYHREWVAE
          HNNDEIGSAPEVLIGYLAASTKIRIKIAGSVMLCHYSYKVAADPFIILSALAPRT
          DLGVKARPGSROLSTDALQAEKKPYROPDELELTLPVRDDPDTTRTIALAPRPO
          VDRKGCITLGGSTESAISAALGISTVFVAFINEEVEVLAEARAFAHLAPGSEAR
          FHLAFAVRAHTKEAEKHAIVRESIKYVLKDKRVNVNVSREDAEAYLENTPEYDII
          YOKTGIIAGTKEEVAEELRLSGTYKINDFVIFPIKNAVEKQLSYQLSDAVLAAR"
          complement(5649. .6428)
          /gene="hisP"
          complement(5649. .6428)
          /gene="hisP"
          /codon_start=1
          /transl_table=11
          /product="histidine transport protein (ATP-binding
          protein)"
          /protein_id="CAB14912.1"
```

```
/db_xref="GI:2635418"
/db_xref="SPTREMBL:O34900"
/translation="WIEIKNLHKQFGJHNLKGINLTVRKEGVTTIIGPSGSGKTPTL
RCLNLERPDDEGIISIHDKVINCSPSKSEVMKOTAMFOQYHLPARKVTIENM
EGLTARKMKODAAVAAYENELRKVGLQDDKINAVPSLSSGSGKORVGIARALHPDV
LIPDEPTALDPELVGEYLEFMLEFIVTGATMIVYTEMEMEARVSDQVVFMDGCVIV
EQGTPEVFRRTKDKDRQFLRRVSPELPEPKHKEPVI"
complement(6425. .7132)
/gene="ytnM"
complement(6425. .7132)
/gene="ytnM"
complement(6425. .7132)
/function="unknown"
/note="similar to amino acid ABC transporter (permease)"
/codon_start=1
/transl_table=11
/protein_id="CAB14913.1"
/db_xref="GI:2635419"
/db_xref="SPTREMBL:O34931"
/translation="MORDEPRIVAMKEVMKTPITLMAVLPVREGFVALGNITVR
IFRKGLVACSRFYVSFFRSPAILHMLIYLGIPIVADOVSSFFHLGMSANEIPVSM
FVIMALSLTAGAVLEIRSGITLMDTGVEAAYSIGLTVSQRVIRILQALKVSIIP
NFTNMGIFLHTTSIAAIVAVPEITGTATVASDNVAFLEAFIGALITVWTLILLES
ANGVLERKAARFOGGTL"
complement(7162. .7881)
/gene="ytnL"
complement(7162. .7881)
/gene="ytnL"
/function="unknown"
/note="similar to amino acid ABC transporter (permease)"
/codon_start=1

alignment_scores:
  Quality: 609.00      Length: 396
  Ratio: 2.316        Gaps: 5
  Percent Similarity: 66.414  Percent Identity: 33.586

alignment_block:
US-09-724-797-36 x BSUB0016/rev ..

Align seg 1/1 to reverse of: BSUB0016 from: 1 to: 215640

      8  AAlpheasPProThrAspaIaaspValaArgaArgasPProTyProSerTy 24
          |||  ::  ::::::::::::::  ::::::::::::::  ||
          91715 GCATCGTCAACGTCATCTCTGAGTTTGAAAAACCCATATCTTTTAA 91666

      24  rHtSTpLeuLeuArgHisAspProValHisArgGlyAla.....H 38
          |||  ::::::::::::::  ::::::::::::::  ::
          91665 CGACACATTGCGAGCTTCATCCATCTATTAAGAGAGAGTTCTTAAT 91616

      38  iSArgValTrpTyValSerArgpheaIaaspValaArgaValaIleuGly 54
          ::  |||  ::::::::::::::  ::  ::::::::::  |||  ::
          91615 ACCCGGCGTGTGATGTACACAGATATGAAGAAACGGCTCTCTATTTGAAA 91566

      55  AspgIuArgpHe...AlaArgThrglyIleArgArgpHeTrpThrasple 70
          |||  |||  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
          91565 GATGCGAGATGTCAAAAGTCCGACCCCGCTGCTGAGAGCTCAACCAATA 91516

      70  vAlaIglyProGlyLeuLeuAlaGluIleValIglyaspIleIleuPheg 87
          |||  ::::::::::::::  ::::::::::::::  ::::::::::  |||
          91515 TCAGGAC.....CTTTCACATGTGCMAAATCAAAAGAGCGTGTTC 91475

      87  lAspIuProAspHisIglyArgLeuArgGlyValaValaIglyProAlaPhe 103
          |||  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
          91474 AGAACAGCCTGATCATATACGATTCGCGACGCTGCCAGCGAGCGTTT 91425

      104  SerProSerAlaIleuArgArgLeuGluProValIleValaglyThrValas 120
          ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
          91424 ACCCGAGAACAGACAGAGATTATCAGCGCTATATCATTTGAAAGTGTCCA 91375

      120  pAspLeuLeuArgProAlaIleuAlaArgGlyAlaIleuAspValaIaspG 137
          |||  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
          91374 TCATTGCTTGATCAAGTGCAGGTAAAAAAGATGAGGTCATCTTGGG 91325
```

[illegible]

274-281  
90234-90257

seq_name:	gb_ba:AF008220
seq_documentation_block:	
LOCUS	AF008220 220060 bp DNA linear BCT 04-FEB-1998
DEFINITION	Bacillus subtilis rrmB-dnaB genomic region.
ACCESSION	AF008220
VERSION	AF008220.1 GI:2293135
KEYWORDS	
SOURCE	Bacillus subtilis.

ORGANISM	REFERENCE
Bacillus subtilis	Bacteria: Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group: Bacillus.	
1 (bases 1 to 7430)	
Green,C.J., Stewart,G.C., Hollis,M.A., Vold,B.S. and Bott,K.F.	
Nucleotide sequence of the Bacillus subtilis ribosomal RNA operon, rrib	
Gene 37 (1-3), 261-266 (1985)	
86031361	
2 (bases 153210 to 153762)	
Comors,M.J., Mason,J.M. and Setlow,P.	
Cloning and nucleotide sequencing of genes for three small, acid-soluble proteins from Bacillus subtilis spores	
J. Bacteriol. 166 (2), 417-425 (1986)	
86195826	
3 (bases 213161 to 218473)	
Ogasawara,N., Moriya,S., Mazza,P.G. and Yoshikawa,H.	
Nucleotide sequence and organization of dnaB gene and neighbouring genes on the Bacillus subtilis chromosome	
Nucleic Acids Res. 14 (24), 9989-9999 (1986)	
87117549	
4 (bases 200404 to 201481)	
Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.	
Cloning and nucleotide sequence of phoP, the regulatory gene for alkaline phosphatase and phosphodiesterase in Bacillus subtilis	
J. Bacteriol. 169 (7), 2913-2916 (1987)	
87250247	
5 (bases 201234 to 203212)	
Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.	
Nucleotide sequence of the Bacillus subtilis phoR gene	
J. Bacteriol. 170 (12), 5935-5938 (1988)	
89053932	
6 (bases 142232 to 144147)	
Grundy,F.J. and Henkin,T.M.	
Cloning and analysis of the Bacillus subtilis rpsd gene, encoding ribosomal protein S4	
J. Bacteriol. 172 (11), 6372-6379 (1990)	
91035248	
7 (bases 133624 to 134990)	
Henkin,T.M., Grundy,F.J., Nicholson,M.L. and Chambliss,G.H.	
Catabolic repression of alpha-amylase gene expression in Bacillus subtilis involves a trans-acting gene product homologous to the Escherichia coli lacI and galP repressors	
Mol. Microbiol. 5 (3), 575-584 (1991)	
91260441	
8 (bases 140810 to 142610)	
Henkin,T.M., Glass,B.L. and Grundy,F.J.	
Analysis of the Bacillus subtilis tyrS gene: conservation of a regulatory sequence in multiple tRNA synthetase genes	
J. Bacteriol. 174 (4), 1299-1306 (1992)	
92138624	
9 (bases 217570 to 220060)	
Putzer,H., Gendron,N. and Grunberg-Manago,M.	
Co-ordinate expression of the two threonyl-tRNA synthetase genes in Bacillus subtilis: control by transcriptional antitermination involving a conserved regulatory sequence	
EMBO J. 11 (8), 3117-3127 (1992)	
92347349	
10 (bases 134990 to 141290)	
Grundy,F.J., Waters,D.A., Yakova,T.Y. and Henkin,T.M.	
Identification of genes involved in utilization of acetate and acetoin in Bacillus subtilis	
Mol. Microbiol. 10 (2), 259-271 (1993)	
95020526	
11 (bases 162129 to 164080)	
Grundy,F.J., Waters,D.A., Allen,S.H. and Henkin,T.M.	
Regulation of the Bacillus subtilis acetate kinase gene by CcpA	
J. Bacteriol. 175 (22), 7348-7355 (1993)	
94042910	
12 (bases 7009 to 15526)	
Kiehl,T.A., Boels,J.M., Beldman,G. and Venema,G.	
Operon encoding enzymes involved in glycogen biosynthesis and	



JOURNAL MEDLINE REFERENCE AUTHORS TITLE	degradation Mol. Microbiol. 11 (1), 203-218 (1994) 94195107 13 (bases 196487 to 200620) Jin,S. and Sonenshein,A.L. Identification of two distinct Bacillus subtilis citrate synthase genes
JOURNAL MEDLINE REFERENCE AUTHORS	J. Bacteriol. 176 (15), 4669-4679 (1994) 94321340 14 (bases 16985 to 19588) Abe,A., Koide,H., Kohn,T. and Matabe,K. A Bacillus subtilis spore coat polypeptide gene, cots
JOURNAL MEDLINE REFERENCE AUTHORS	Microbiology 141 (Pt 6), 1433-1442 (1995) 95400496 15 (bases 131934 to 133970) Bolotin,A., Khazak,V., Stoyanova,N., Ratmanova,K., Yomantas,Y. and Kozlov,Y. Identical amino acid sequence of the aroA(G) gene products of Bacillus subtilis 168 and B. subtilis Marburg strain
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	Microbiology 141 (Pt 9), 2219-2222 (1995) 96118703 16 (bases 25258 to 31212) Rowland,B., Hill,K., Miller,P., Driscoll,J. and Taber,H. Structural organization of a Bacillus subtilis operon encoding menaquinone biosynthetic enzymes
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	Gene 167 (1-2), 105-109 (1995) 96144257 17 (bases 196487 to 200620) Jin,S., De Jesus-Berrios,M. and Sonenshein,A.L. A Bacillus subtilis malate dehydrogenase gene
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	J. Bacteriol. 178 (2), 560-563 (1996) 96134995 18 (bases 129888 to 132207) Varon,D., Brody,M.S. and Price,C.W. Bacillus subtilis operon under the dual control of the general stress transcription factor sigma B and the sporulation transcription factor sigma H
JOURNAL MEDLINE REFERENCE AUTHORS	Mol. Microbiol. 20 (2), 339-350 (1996) 96310371 19 (bases 81540 to 91690) Bower,S., Perkins,J.B., Vocum,R.R., Howitt,C.L., Rahaim,P. and Pero,J. Cloning, sequencing, and characterization of the Bacillus subtilis biotin biosynthetic operon
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	J. Bacteriol. 178 (14), 4122-4130 (1996) 96312354 20 (bases 49093 to 51682) Vocum,R.R., Perkins,J.B., Howitt,C.L. and Pero,J. Cloning and characterization of the metE gene encoding S-adenosylmethionine synthetase from Bacillus subtilis
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	J. Bacteriol. 178 (15), 4604-4610 (1996) 96345628 21 (bases 100760 to 102298) Kappes,R.M., Kempf,B. and Bremer,E. Three transport systems for the osmoprotectant glycine betaine operate in Bacillus subtilis: Characterization of Opud
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	J. Bacteriol. 178 (17), 5071-5079 (1996) 96393937 22 (bases 1 to 220060) Lapidus,A., Galleron,N., Sorokin,A. and Ehrlich,S.D. Sequencing and functional annotation of the Bacillus subtilis genes in the 200 kb rnmB-dnB region
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	Microbiology 143 (Pt 11), 3431-3441 (1997) 96048467 23 (bases 1 to 220060) Lapidus,A., Galleron,N., Sorokin,A. and Ehrlich,D. Direct Submission
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	Submitted (13-JUN-1997) Laboratoire de Genetique Microbienne, INRA, Domaine de Vilvert, Jouy-en-Josas cedex 78352, France
FEATURES	Location/Qualifiers
Source	1..220060
	/organism="Bacillus subtilis"
	/db_xref="taxon:1423"

[illegible]

```

88862 GATCGGAGATTCAAAAGTCCGCGACCCCGCTGCTGAGAGCTCAACCAATA 88911
70 uvalGlyProGlyLeuLeuAgluIleValGlyAspIleIleuPheG 87
88912 TCAGGAC.....CTTTCATGTGTGCAAAATCAATGATGCTGTTC 88952
87 InaSerGluProAspHisGlyArgLeuArgGlyValValGlyProAlaPhe 103
88953 AGAACGACCGCATCATAGAGATGGGAGAGCGCTGGCCAGCGAGCGCTT 89002
104 SerProSerAlaLeuArgArgLeuGluProValIleAlaGlyThrValAs 120
89003 ACCGCGGAAGACGACAGAGATTATCACCGCTATTCATTGCAATGTCACA 89052
120 PASLeuLeuArgProAlaLeuAlaArgGlyAlaMetAspValValAspG 137
89053 TCATTGTGCTTATCATAGTCAGAGTAAAAAAAAGATGGAGTCTTTCCG 89102
137 LuLeuAlaATyrProLeuAlaLeuArgAlaValLeuGlyLeuGlyLeu 153
89103 ACTTGCCTTTCTTTAGCAAGTTTGTGCATAGCTAACATATATAGTGTA 89152
154 ProAlaAlaAspTrrpGlyAlaValGlyArgTrrpSerArgAspValGlyAr 170
89153 CCGGAGGAGATGAGGAGCAATTAAAGAGTGGCTCGAGTCTCATTTCA 89202
170 gThrLeuAspArgGlyAlaSerAlaGluAspMetArgArgGlyHisAla 187
89203 AACCATTTATTTACCCGCTCAAGAAAGCATTAACAGAGGCAATTTA 89252
187 laIleAlaGluPheAlaAspTyrValGluArgAlaLeuAlaArgArg 203
89253 TGGCGTGCAGGCTATGCGCATTTCAAGAGTCAATTCAAAAGAGAAA 89302
204 ArgGluGlyGlyGluAspLeuLeuAlaLeuMetLeuAspAlaHisAspAr 220
89303 CGCCACCTCTACACAGATATGATGCGCTCTGTGAAGGGAGAGAAA 89352
220 gGlyLeuMetSerArgAsnGluIleValSerThrValValThrPheIleP 237
89353 GGATTAAGCTACGGAAGAGAGGCGGATCTACAGCATATGCTGGGGA 89402
237 heThrGlyHisGluThrValAlaSerGluValGlyAsnAlaValLeuSer 253
89403 TCGCGGACATGAGACAAAGGTCATCTCATACCAATTCAGTCTTGT 89452
254 LeuLeuAlaHisProAspGluLeuAspLeuLeuArgArgProAspIle 270
89453 CTGCTGCAGCATCCGAAACAGCTTTTGAACCTGAGAGAAATCCAGATCT 89502
270 uLeuAlaGluAlaValGluGluCysLeuArgTyrAspProSerValGlns 287
89503 TATTGGTACCGGAGTGCAGGATGTTTACGCTATGAAAGCCCAACGAAA 89552
287 eTAsnThrArgGluLeuAspValAspValGluLeuArgGlyArgArgLeu 303
89553 TGACGCGCAGAGTTGCTGCAGAGATATGACATCTGGCGGTGACGATC 89602
304 ArgArgAspAspValValValValLeuAlaGlyValAlaAlaAsnArgAspR 320
89603 CGTCAAGGAGAAACAGCTATCTTTTGTAGAGCGGCTAATCGAGACC 89652
320 cArgArgTyrAspArgProAspAspPheAspIleGluArgAspProValP 337
89653 TAGCATATTTACGAAGCCCGATGCTCTGATATTAACGAGAGTCTCTATC 89702
337 roSerMetSerPheGlyAlaGlyMetArgTyrCysLeuGlySerTyrIleu 353
89703 CGCATCTTTTCATTTGCGGCATGCGCATCATGTTTGTAGGGTCTCGCTG 89752
354 AlaArgThrGlnLeuArgAlaAlaVal...AlaAlaLeuAlaArgLeuP 369
89753 GCACGATTAGAGCGCAAAATTCGATTAACTCTTTCGACGAGAAATGCC 89802

```

```

369 oGlyLeuArgLeuGlyCysAlaSerAspAlaLeuAlaTyrGlnProArgT 386
1:::|||||  |||:|
89803 CAGCCTTATCTT.....GCGGATTTTGAATGGCGGTATCGCGCGCTT 89846
386 hrMetPheArgGlyLeuAlaSerLeuProIleAlaPhe 398
|||||  |||:|
89847 TTGATTTTCGCGCGCTTGAGAGAGCTGCGGTGATCTTT 89884

```

seq\_name: gb\_ba:SCH10

seq\_documentation\_block:

LOCUS SCH10 39524 bp DNA linear BCT 04-MAY-1999  
 DEFINITION Streptomyces coelicolor cosmid H10.  
 ACCESSION AL049754  
 VERSION AL049754.1 GI:4753846  
 KEYWORDS ABC transporter; adenylsuccinate synthetase; aspartate  
 aminotransferase; ATP-dependent protease ATP-binding subunit;  
 cytochrome P-450 hydroxylase; fba; fructose 1,6-bisphosphate  
 aldolase; lipase/esterase; oxidoreductase; purA; RNA polymerase  
 sigma factor; spermidine synthase; transmembrane efflux protein;  
 two-component regulator; two-component sensor kinase; uridine  
 5'-monophosphate synthase.

#### SOURCE

##### ORGANISM

Streptomyces coelicolor A3(2).  
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 1 (bases 1 to 39524)  
 Redenbach, M., Kleiser, H.M., Denaplatte, D., Eichner, A., Cullum, J.,  
 Kinashi, H. and Hopwood, D.A.  
 A set of ordered cosmids and a detailed genetic and physical map  
 for the 8 Mb Streptomyces coelicolor A3(2) chromosome  
 Mol. Microbiol. 21 (1), 77-96 (1996)  
 97000351

#### TITLE

##### REFERENCE

##### AUTHORS

##### REFERENCE

##### JOURNAL

##### AUTHORS

##### TITLE

##### JOURNAL

##### COMMENT

##### REFERENCE

##### AUTHORS

##### REFERENCE

##### JOURNAL

##### AUTHORS

##### REFERENCE

##### JOURNAL

##### AUTHORS

##### REFERENCE

##### JOURNAL

##### AUTHORS

##### REFERENCE

##### JOURNAL

##### AUTHORS

##### REFERENCE

##### JOURNAL

##### AUTHORS

Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC.  
 Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.  
 (URL: [http://www.sanger.ac.uk/Projects/S\\_coelicolor/](http://www.sanger.ac.uk/Projects/S_coelicolor/))  
 CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).  
 The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.  
 The length in codons is given for each CDS.  
 Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.  
 The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.  
 Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov/jb/>  
 jun/cgl-bin/frameplot.pl. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (atc)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.  
 IMPORTANT: This sequence MAY NOT be the entire insert of the

sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid H10 lies between 66T3 and H44 on the AseI-H genomic restriction fragment.

## FEATURES

## Source

1..39524  
/organism="Streptomyces coelicolor A3(2)"  
/strain="A3(2)"  
/db\_xref="taxon:100226"

## misc\_feature

1..44  
/gene="SCH10.01"  
/note="Overlap with S.coelicolor cosmid St66T3 from 35058 to 35101"

## CDS

/gene="SCH10.01"  
/note="SCH10.01, hypothetical protein, partial CDS, len: >119 aa; probable CDS suggested by positional base preference, GC frame analysis and amino acid composition"  
/codon\_start=1  
/transl\_table=1  
/label="SCH10.01"  
/product="hypothetical protein"  
/protein\_id="CAB42010.1"  
/db\_xref="GI:4753847"  
/translation="DPRMAAGVRRGEGEQAQAAGDITVIGDYRRELATLFGNPLCDFLHRLVQSWCTVCHLRLSELKALMSGHALVDALARRDVGARALVDAYNSHSLALEGLAGE"  
1..360  
/gene="SCH10.01"  
462..1127  
/gene="SCH10.02"  
462..1127  
/gene="SCH10.02"  
/note="SCH10.02, hypothetical protein, len: 221 aa; unknown function, probable CDS suggested by positional base preference, GC frame analysis and amino acid composition"  
/codon\_start=1  
/transl\_table=1  
/label="SCH10.02"  
/product="hypothetical protein"  
/protein\_id="CAB42011.1"  
/db\_xref="GI:4753848"

## gene

## gene

## CDS

/translation="MRSPLACDLMLVLYVLCPPNPFAELAOYDVLAEGIPVYVQVPGLSGEVAPAGFDVPAHLRAHLIYCGLPVTVDELGDGYOLEMFEATAQOSHLYWYHDHAGAYPVVFPPLPSSDELLAGGPGSSHTLIRLEVAALGIDPANPAPPPPLAPPELEBPVAPAHDSPPARERHWTGLHAAATRSIAQSMIVFS"  
complement(1143..1433)  
/gene="SCH10.03c"  
complement(1143..1433)  
/gene="SCH10.03c"  
/note="SCH10.03c, hypothetical protein, len: 96 aa; unknown function, probable CDS suggested by positional base preference, GC frame analysis and amino acid composition"  
/codon\_start=1  
/transl\_table=1  
/label="SCH10.03c"  
/product="hypothetical protein"  
/protein\_id="CAB42012.1"  
/db\_xref="GI:4753849"

## gene

## CDS

/translation="WRVLLKATLDTEKSNEVINGRMEALIKETVDRIGEAAYFSTIDGRRFALLVFDMDSSQMPPIGEPELMELNAEVDLCPIMNGDDLAKGLSGLG"  
complement(1472..2356)  
/gene="SCH10.04c"  
complement(1472..2356)  
/gene="SCH10.04c"  
/note="SCH10.04c, probable RNA polymerase sigma factor, len: 294 aa; similar to many e.g. TR-005735 (EMBL:U08308), stige, Mycobacterium avium extracytoplasmic function

## gene

## CDS

alternative sigma factor (251 aa), fasta scores, opt: 226  
z-score: 266.2 E(): 1.6e-07, 29.8% identity in 168 aa overlap"

/codon\_start=1  
/transl\_table=1  
/label="SCH10.04c"  
/product="putative RNA polymerase sigma factor"  
/protein\_id="CAB42013.1"  
/db\_xref="GI:4753850"

/translation="MPARSDMARQEFELDRPRLRAYAYRILSGEADDAIOEAMLRA  
DRANTSEVGNPSGMLTIVAVCLNLRAADTRREPELDAAAROPAAATGAADPAE  
VOLADEVIGIALITVDTLGPABERLATVTHMPEPDDIAAMLDKPAATROLASRAR  
RRVRCVAPPENDIPRRRRAYDAITLATRGDFPALVALLHPVLSADAAVPTPEPV  
TVSGMEGVARAAASAKARSAVAIVLDGVGMAENGRLRVYLFKPDRAUGRIITGI  
DVIAEPARLNELDITGIG"  
complement(2472..2672)  
/gene="SCH10.05c"  
complement(2472..2672)  
/gene="SCH10.05c"  
/note="SCH10.05c, hypothetical protein, len: 66 aa; unknown function, probable CDS suggested by positional base preference and GC frame analysis"

## gene

## CDS

/codon\_start=1  
/transl\_table=1  
/label="SCH10.05c"  
/product="hypothetical protein"  
/protein\_id="CAB42014.1"  
/db\_xref="GI:4753851"  
/translation="MTDRYSASACQGGYGGNGPEDCGDPVREVARHLRAPLRCV  
HIGPSLILATGVLMWPCGILVR"  
complement(2749..3240)  
/gene="SCH10.06c"  
complement(2749..3240)  
/gene="SCH10.06c"  
/note="SCH10.06c, hypothetical protein, len: 163 aa; unknown function, probable CDS suggested by positional base preference, GC frame analysis and amino acid composition"  
/codon\_start=1  
/transl\_table=1  
/label="SCH10.06c"  
/product="hypothetical protein"  
/protein\_id="CAB42015.1"  
/db\_xref="GI:4753852"

## gene

## CDS

/translation="MRKPTRTWTWLLAAVVVVTGAFYVDGHRVSAPAHAGPKCDE  
IVSRLPDIPGASRDMALGDGVSMGOKAVFGCAEELQPNVNLCTVADGVWDVDE  
ARLRDGVSVLRITGRSPAYEFTYSGPREVGILVALDPVAVWIPQEKRCIGLDAA  
TVL"  
complement(3412..4695)  
/gene="SCH10.07c"  
/note="SCH10.07c"  
/note="pura"  
complement(3412..4695)  
/gene="SCH10.07c"  
/note="SCH10.07c"  
/note="SCH10.07c, pura, probable adenylosuccinate synthetase, len: 426 aa; similar to many e.g. SW:PURA.ECOLI (EMBL:J04199), pura, Escherichia coli adenylosuccinate synthetase (431 aa), fasta scores, opt: 1421 z-score: 1608.0 E(): 0, 50.5% identity in 420 aa overlap. Contains Pfam match to entry PF00709 Adenylosucc-synt, Adenylosuccinate synthetase, score 742.40, E-value 1.9e-219, PS00513 Adenylosuccinate synthetase active site and PS01266 Adenylosuccinate synthetase GTP-binding site"  
/codon\_start=1  
/transl\_table=1  
/label="pura"

## gene

## CDS

alignment\_scores: 572.50 Length: 413  
Quality: 2.272 Gaps: 8  
Ratio: 61.017 Percent Identity: 33.656  
Percent Similarity: alignment\_block:





CDS

AAVGIKEDGAVIVRPGFVGWRTATTRNAVETSTQHLRDVLDRIILROKA"

15034. .16253  
/codon\_start=1  
/transl\_table=1  
/product="cytochrome P450"  
/protein\_id="BAB69286.1"  
/db\_xref="GI:15824120"  
/translation="WDKADADYPRYAMLRAREPVRQYLEGGTGLLTATREYDKRL  
SDPRMSKDPNAPLDMDQAGKRPLEDRTGLTLLTTDAFENRLRLKLVSTATTAR  
VEGLRAQVQHTDGLDITVRGQAEILGDAFPLATVIGELGVPRADQVFRWT  
KDFRWMTDSAOADRDGAPVGLRDLEYLRVLDKRRDPADGLVDALIAADDD  
RLNEAELISMMSLLVGFETVNLIGNGTLLALRPDQLLRERBELVDSALEML  
RYDSFEETATRPLEPIEVAGTRIEKGPIVLLASNNRDGAFKFPAPDDVDYRADP  
AHVARGAHFICLAPRLARLEGRILAFHLLRLRGLALSVPEQLRMGRSLTVGLER  
LPVTFDA"

## alignment\_scores:

Quality: 563.00 Length: 406  
Ratio: 2.298 Gaps: 9  
Percent Similarity: 60.345 Percent Identity: 33.498

## alignment\_block:

US-09-724-797-36 x AB070947 ..

Align seg 1/1 to: AB070947 from: 1 to: 24268

15 AspyValrArgarAspProtyrProserTyrHisTrpleuLeuargHisas 31  
|||||  
15057 GATTTCCGGCGCTGATCCCTATCCGAGTACGGGTACGTACGTCCGGAGGA 15106  
31 pProvalHis.....ArgGLyAlaHisArgValTrpTyv 43  
|||||  
15107 ACCGGTCCGGAGGCTCTGGAGGGCCCGGCTTTACGGACTG...CTGG 15153  
43 aLSerArgpheaAlaAspValaArgAlaValleuGLyAspGLuArgpheaAla 59  
|||||  
15154 TGACCCGGTACGAGACGTGGCGAAGCTGCTGCATCCCGGATGAGC 15203  
60 Arg.....ThrGLyLeuArgar 65  
|||||  
15204 AAGGACCCGGCGAAGCAGCGCTGACTGGCAGAGGGCGGCAAGGGAG 15253  
65 gPheTrpThrAspLeuValGLyProGLyLeuLeuAlaGLuILeValGLyA 82  
|||||  
15254 ACCGCTGGAGAGCCGAGACGGG.....CTCGGGA 15282  
82 spILleLeuPheGLAspLeuProAspHisGLyArgLeuArgGLyVal 98  
|||||  
15283 CGCATCTGCTGACCCAGCGACGCGCCGACACACACGCGGCTGCGCTG 15332  
99 ValGLyProAlaPheSerProserAlaLeuArgTrpLeuGLuProVal11 115  
|||||  
15333 GTCTGACCGGCTTACCGCCCGCGGCTGAGGGCTGCGCGACAGGT 15382  
115 eaGLyThrValAspAspLeuArgProAlaLeuAlaArgGLyAla1w 132  
|||||  
15383 GCAGACATCAACGACGGGCTCTCGACACATGTCGCCCGCGGAGG 15432  
132 etaSpValValAspGLuLeuAlaTrpProLeuAlaLeuArgAlaValleu 148  
|||||  
15433 CCGACTGATGCGGACTTCCGTTCCCGTCCGATCAGCGTGAATCTGT 15482  
149 GLyLeuGLuGLyLeuProAlaAlaAspTrpGLyAlaValGLyArgTrpSe 165  
|||||  
15483 GAACGTGCGGAGTCCCAAGCGGACAGACAGTGTCCCGCCAGTGAGAC 15532  
165 rArgspValGLyArgTrpLeuAspArgGLyAlaSerAlaGLuAspMetA 182  
|||||  
15533 GAAGACATTCGGCGGTGGAGCAACACGACTCCGCCAGCGCGAC... 15578  
182 rArgArgGLyHisAla.....AlaILleAlaGLuPheAlaAspTyVal 195  
|||||  
15579 ..CGCGTGAACGCGCGGCTGGACTGCGGACCTCTCGAATATCTG 15626

196 GIUArgAlaLeuAlaArgArgArgGLuGLuGLyGLuAspLeuLeu.. 211  
|||||  
15627 ACCGCACTGGTGCAGACAGCGCCGTACAGACCCCGCGGACGACTGCTGA 15676  
212 AlaLeuMetLeuAspAlaHisAspArgGLyLeuMetSerArgansGLuI 228  
|||||  
15677 CGCGGTGATCCGGCGCGCGGAGCAGACGACCGGCTGCAAGACGCGCGAGT 15726  
228 LeValSerThrValValThrPheILlePheTrpGLuHisGLuThrValAla 244  
|||||  
15727 TGCCTGCATGATGTCCTGCTGCTGGCGGCGGCTTGCAGACAGGCTC 15776  
245 SerGLuValGLyAspAlaValLeuSerLeuLeuAlaHisProAspGLuLe 261  
|||||  
15777 AATCTGATCGGACAGCGCACCTGGCCCTGCTGGCGCACCCCGACCACT 15826  
261 uAspLeuLeuArgArgArgProAspLeuLeuAlaGLuAlaValGLuGLu 278  
|||||  
15827 CCCTGCTGCGCGAGCGACCGCACTGCTGACTGCGGCTGGAGGAGA 15876  
278 yLeuArgTrpAspProserValGLuSerAspThrArgGLuLeuAspVal 294  
|||||  
15877 TGCCTGCGTACGACGGGTCTGTCGAGACGGCGAGCTGGCGGTTCCGCTC 15926  
295 Asp...ValGLuLeuArgGLyArgArgLeuArgArgAspAspValVala 310  
|||||  
15927 GAACCATGAGAGTGGCGGCGACGGCGCATCGAAGAGGCGCACCCGCTCT 15976  
310 lValleuAlaGLyAlaAlaAsnArgAspProArgArgTrpAspArgProA 327  
|||||  
15977 CCTGTCCTCGCATGCGCCACCGGGAGCGGGCGAAGTTCGCGGACCGG 16026  
327 spAspPheAspILleGLuArgAspProValProserMetSerPheGLyAla 343  
|||||  
16027 ACGACTTCACAGTCACCCCGCGGACCCCGCCACAGTGGCTTCGGGGG 16076  
344 GLyMetArgTrpCysLeuGLySerTrpLeuAlaArgTrpGLuLeuArgAl 360  
|||||  
16077 GGTGCGCACTTCTGCTGCTGCTCCCTGGCCGCGGTGAGGCGCGGAT 16126  
360 aAlaVal...AlaAlaLeuAlaArgLeuProGLyLeuArgGLyCysA 376  
|||||  
16127 CGCGTTCACAGCATGCTCGCGCGGCTGCGCGCTTGGCTGCTCGTAC 16176  
376 lAserAspAlaLeuAlaTrpGLuProArgTrpMetPheArgGLyLeuAla 392  
|||||  
16177 CGCGGAGCAGATTGCGATGCGAGCGAGCTCAGCGTCCGAGACTGGA 16226  
393 SerLeuProILleAlaPhe 398  
|||||  
16227 GCCCTCGCGTCACTTTC 16244  
seq\_name: gb\_ba:AE004675  
seq\_documentation\_block:  
LOCUS AE004675 10060 bp DNA linear BCT 30-AUG-2000  
DEFINITION Pseudomonas aeruginosa PA01, section 236 of 529 of the complete genome.  
ACCESSION AE004675 AE004091  
VERSION AE004675.1 GI:9948522  
KEYWORDS  
SOURCE Pseudomonas aeruginosa.  
ORGANISM Pseudomonas aeruginosa  
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
Pseudomonas.  
REFERENCE  
1 (bases 1 to 10060)  
AUTHORS Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrenner,P., Hickey,M.J., Brinkman,F.S., Huftagle,W.O., Kowalik,D.J., Lagrou,M., Garber,L.J., Goltry,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R., Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.  
TITLE Complete genome sequence of Pseudomonas aeruginosa PA01, an

JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
2 (bases 1 to 10060)  
Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warrenner P., Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M., Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y., Brody L.L., Coulter S.N., Folger K.R., Kas A., Iarbig K., Lam R.M., Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T., Reitzer J., Sailer M.H., Hancock R.E.W., Lory S. and Olson M.V.  
TITLE  
JOURNAL  
Direct Submission  
Submitted (16-MAY-2000) Department of Medicine and Genetics,  
University of Washington Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA  
FEATURES  
Location/Qualifiers  
1..10060  
/organism="Pseudomonas aeruginosa"  
/strain="PA01"  
/db\_xref="taxon:287"  
70..1416  
/gene="PA2472"  
70..1416  
/gene="PA2472"  
/codon\_start=1  
/transl\_table=11  
/product="probable MFS transporter"  
/protein\_id="AAG05860.1"  
/db\_xref="GI:9948523"  
/translation="MPTPDKNKKIVQVAFIDAQAVSAFQRLVSLICELIIVAMGFDITA  
AIGFIAPALAHMDWSPALSPILGALAGALGAFAGAPLADRGKRSVLLSYLF  
GGMSLASAVGSEVETLALREFTGILGAMPNATITLSECPRRHRLAMTYAMCGF  
TLGSAIGLILARMPALGMEVILLGGPLASPLLMACLPESVREFLRAPDAAR  
LRIITIERGLRDPDWAGRIELPADSEDSPLRIOLIGALRGSTILLMATPEMGLIIT  
VLTITWMLPTLIGTGSGESLGEATYISAMOLGTCGAILLGSAMDFFDHRRLSLAYV  
GALFTLGLASLHSPALLALCYAGVGFTSGSOGANKLADFTFRSRATGVSNALG  
LGRGISVGSLSGSGALLGLGSGSGLLALVLPALLAAVAHRLGRRRASPPTL"  
1429..2073  
/gene="PA2473"  
1429..2073  
/gene="PA2473"  
/codon\_start=1  
/transl\_table=11  
/product="probable glutathione S-transferase"  
/protein\_id="AAG05861.1"  
/db\_xref="GI:9948524"  
/translation="MOLYSFENSSTSYVRITALKGIQYOVVYNLRQSGRLRPADK  
QRNPMGALPTIVDADGRRSOSLAIITVIDAVQPPRLIPDPLHRAQALELILVAC  
DIHPLANNRVAKYLTIVUGLIDAEEDRQRMVAHWAVGLAAETLLNRHRRGAFPAQAAA  
GIVECCLVPEQLANARRMCGDLAPYPALLEGRCLAEAFORASPEROPDYLPP"  
2115..3026  
/gene="PA2474"  
2115..3026  
/gene="PA2474"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAG05862.1"  
/db\_xref="GI:9948525"  
/translation="KTRTRRASLLALPLLPSFOAILPQCNLQGTSTFDGKALPGT  
LHQQFVTEADDFDANGHRRLDNRKTLTATSHFAHISOHLFGAYVAAEILLVY  
LLADDFEPHAGSANGNLIVSPILQMPQTELEPRWQRLANPLADPTGRYDRHR  
AINPGSNHSEFNPHYALTFWEFAGWEASARLOYLPGKNRPAPPLADYLOPGGAFH  
ANYSVSLADAHMRGLSGYSLROYTDERVGHDRPDSRERDAIGPALMGWGKSRLL  
PLNAYHEGARNRSEGRLLIIRYSQVF"  
complement(3008..4342)  
/gene="PA2475"  
complement(3008..4342)  
/gene="PA2475"  
/codon\_start=1  
/transl\_table=11  
/product="probable cytochrome P450"  
/protein\_id="AAG05863.1"

/db\_xref="GI:9948526"  
/translation="MDDAFSEEGSAOPRHDAOPALAPRSGDFIHTYHPDFVADPEP  
LRLIRSRAPVCRDQASIMWISRYADVSACLRDRRSADPARIGAGVCGGASMGCH  
QOQOLPARKFDNREPLFNDAPRRHRLRLPAPAGPPVAVRMERITVYBELDLSLE  
RREDFLRLDREPLTTRVAEELGFPREDYGLLPWRDLAAGDLAASBGDQINR  
SAVAFSYDLROARGMSDSSRPPSGAASPLDGAAMLENGLEDLVAAYAAVFMAA  
FEETISVAGNATLALHPDQLDLRCPELAAVBEILFEGVARGVCITLVEVE  
IGGRIIPGGKVMVLSFLAANDREMFAPDRILOLRANKOHVAFAGPHVCIGAYLA  
RLQOCLRLQVRRRLASEPDLRRRSVRITLERLPYVEGDAOKTCE"  
complement(4415..5185)  
/gene="dsbg"  
/note="PA2476"  
complement(4415..5185)  
/gene="dsbg"  
/codon\_start=1  
/transl\_table=11  
/product="thiol:disulfide interchange protein DsbG"  
/protein\_id="AAG05864.1"  
/db\_xref="GI:9948527"  
/translation="MRLPNLTILGLGLTLNNGLATVAEELPAPIRMYOEKARILG  
SFADPDGLRGYAAEYQOMGALVLTTPGKHVTLGHLFDAGKDLSPRELERLYAPLA  
KEMQKMEQSAWADGRADAARYVYLFSDFNCGYCMFMEQARPWDAQVQURLHVM  
GITREDESKASAILASKDPQKALHDEQCKASTLKPLAKITPAAYRKQLAGMELME  
SMGAATPAIFYLNAEGRMQOQGAQPDQLAEILGPR"  
complement(5182..6018)  
/gene="PA2477"  
complement(5182..6018)  
/gene="PA2477"  
/codon\_start=1  
/transl\_table=11  
/product="probable thiol:disulfide interchange protein"  
/protein\_id="AAG05865.1"  
/db\_xref="GI:9948528"  
/translation="MISFSISGLPYVSIQHLTLFLAIGALLGGMVAARGGRAPVYL  
FNLLILULIARLAFVRYRQYRQGRDPAQMLDLDROGFLAMPPLILAVGALFWMNR  
PALRSIGVASUGLAWLIGSLGILYERGTLPPELISLNAAGESVQIADFGRLV  
INWASCPPCREMPVLOQAQENPDVYELFANQGSATVHFLQGEMLRIDLNLLF  
DNGQIQOQVGSVALPTVFYTAEGRLGSHLGLSRLARYLEAFEPMAAPATRS  
SE"  
complement(6018..7781)  
/gene="PA2478"  
complement(6018..7781)  
/gene="PA2478"  
/codon\_start=1  
/transl\_table=11  
/product="probable thiol:disulfide interchange protein"  
/protein\_id="AAG05866.1"  
/db\_xref="GI:9948529"  
/translation="MRYLILLMLLPGLSQAOPGDDLFAPRGATQTDLPVEKARFE  
TWERLDDGOVQLRWQIAPGYLYQKRLRFPGDLPALQPOLPPESSHDEFEGSYQR  
QSELTLPAAAGQRLGMOGCADAGCTYPPQSOALDGGTGAAGTSGEVAEDGYL  
AGSLQAGNLAMSLILPFGILLAPACSLPMPITLIGLVGSGAGRRRLLAGSYV  
LSMALIYAGLCYVAALLGSLQWLQOQPMILGSRFALFVFLALPMGFEPLQIPALR  
DRDGLSRKRGKSLACAAALGALSGLLVPCMTAPLACALITYAOTGNALHGGVLYF  
SLGIGIGMPLLLLVTVGSRFLPKRGPMMNVKGVFGFLGTAMILLRLPLGLGALMG  
LGGALLLVLYAAVLAHTARGLARHVAVLEGAAGCTFGMGAMLLGAAAGADDPKRPQY  
VAAANRGATPTASHEAFLTVSQPAEILROLAAKAGQVLLDYADMCVSCRIMEK  
OVFAKPDVLAALQGVRLRLDVTADNAASSELLHRYQVPEPSSLIMIGPEGERRARR  
LTGCVADAGCFLAHQATRERG"  
7944..8624  
/gene="PA2479"  
7944..8624  
/gene="PA2479"  
/codon\_start=1  
/transl\_table=11  
/product="probable two-component response regulator"  
/protein\_id="AAG05867.1"  
/db\_xref="GI:9948530"  
/translation="MHVLTREDDDLTASGIVAGLNAOGLTVDRVASAADTQALLOYAR  
FDVLVYLDGLPDEGLRLRLORLROQGVDPVYVLTARADATDRAVLQAGADQYLLK  
FDURELGARLHTLQRRSAGRCVAVTEHGRSLSTPSTRETYLDSRPVELSRRBALILOA  
LLNNGRILISGEOLKDSYVGFGEVESNALNVHIIHLRRLGNAIVQTVKGLGYRLGP  
LRGDGDA"





```

gene      122. 1372
          /gene="pick"
          122.1372
CDS
          /gene="pick"
          /function="synthesis of piromycin, methymycin, and
          neomethymycin"
          /codon_start=1
          /transl_table=11
          /product="cytochrome P450 monooxygenase"
          /protein_id="AAC64105.1"
          /db_xref="GI:3746884"
          /translation="MRTQGTASPPVLDLGLGQFNADPYPTYARLAECPAHNRV
          LPIKGEDEWLAVGVDRAVALADPREFSKMWRNSTPLTEAALNHNMLSEDPHRRH
          LRKIVAEFTFMRVREILRPVQELIVDGLVDMALAPDGRADLMELPLITYISTE
          LGVEPDEAARFVWTFVFPDDPAQAGNANMSEGYSLRLDSRGQDGDLLSALY
          RPSDEGSRLLSEILGMAHLLVAGHETTVNLINGRMLLSHPDQALAKMDLTL
          DGAYPEMLRIKRGPEVSAKTRFVEVPVDLGGTVLPAGDTVLVVLADAHRTPEFPDPRH
          FPIKLDVGHLAGHGHGHCIGAPLARLARLARVALRLERCPDLADVSPSELVWYPNP
          FMIRGKALPLRMWRGRGEAGRRG"
BASE COUNT      208 a      559 c      480 g      223 t
BRIGIN

```

Alignment_scores:	Quality:	538.50	Length:	398
Ratio:	2.216		Gaps:	8
Percent Similarity:	61.055		Percent Identity:	34.673
alignment_block:				
US-09-724-797-36 x AF087022 ..				
Align seg 1/1	to: AF087022	from: 1	to: 1470	
15	AspValaArgArgAspProTyrProSerTyrHisIstriPleuLeuAArgHis	31		
188	GATTTCGCGCGCGATCGCTATCCGACGATCCGAGACACTGCGTCCGAGGG	237		
31	ProValHisArg.....GlyAlaHisArgValTrrPtyrValS	44		
238	TCCGCGCCACCGGGGTCGCGACCCCGAGGGGACGAGGTGGTGCGTGC	287		
44	eArgPheAlaAspValaArgAlaValaLeuGlyAspGlyArgPheAlaArg	60		
:::				
288	TCGGCTACGACCGCGCGGGGCGTCTCGCGAGTCCCGGTTTAGCAGAC	337		
61	ThnGlyIleArgArgPheTrrPthrAspLeuValaGlyProGlyLeuLeuAl	77		
	:::		:::	
338	GAC.....TGGCGCAACTCCACGACTCCCTCGACCGAGGC	372		
77	aglu...IleValaGlyAspIleIleLeuPheGlaAspLupProSphS	93		
373	CGAGGCGCGGCTCAACCAACATCTGAGTCCGACCCGCGGGGACCA	422		
93	IyArgLeuArgGlyValaGlyProAlaPheSerProSerAlaLeuArg	109		
423	CCGCGTGGCGCAAGCTGTGGTGGCCCTGTAGTTCACCATGCGCGGGTGCAG	472		

[illegible]

seq\_name: ggb\_ba:AF079139

## seq. documentation block:

LOCUS AF079139 4342 bp DNA linear BCT 28-OCT-1998  
 DEFINITION Streptomyces venezuelae pikC operon, complete sequence.  
 ACCESSION AF079139  
 VERSION AF079139.1 GI:3800839  
 KEYWORDS Streptomyces venezuelae.  
 SOURCE Streptomyces venezuelae.  
 ORGANISM Streptomyces venezuelae.  
 Bacteria: Firmicutes; Actinobacteria; Actinobacteridae.  
 Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 REFERENCE 1 (bases 1 to 4342)  
 AUTHORS Xue, Y., Zhao, L., Liu, H., and Sherman, D.H.  
 TITLE A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuelae: architecture of metabolic diversity  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (21), 12111-12116 (1998)  
 MEDLINE 9844533  
 REFERENCE 2 (bases 1 to 4342)  
 AUTHORS Xue, Y., Wilson, D., Zhao, L., Liu, H., and Sherman, D.H.  
 TITLE Hydroxylation of macrolactones YC-17 and narbomycin is mediated by the pikC-encoded cytochrome P450 in Streptomyces venezuelae  
 JOURNAL Chem. Biol. 5 (11), 661-667 (1998)  
 MEDLINE 99051447  
 REFERENCE 3 (bases 1 to 4342)  
 AUTHORS Xue, Y., Wilson, D., and Sherman, D.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-JUL-1998) Department of Microbiology, University, 420 Delaware Street SE 1060, P.O. Box 196, Minneapolis, MN 55455, USA  
 COMMENT On Oct 28, 1998 this sequence version replaced gi:377565.  
 FEATURES  
 Source location/Qualifiers  
 1..4342  
 /organism="Streptomyces venezuelae"  
 /strain="ATCC15439"  
 /db\_xref="ATCC:15439"  
 /db\_xref="taxon:54571"  
 122..4291  
 /gene="pikC operon"  
 /note="Involved in the biosynthesis of macrolide antibiotics methymycin, neomethymycin, and pikromycin"  
 122..1372  
 /gene="pikC"  
 122..1372  
 /gene="pikC"  
 /function="catalyzes the hydroxylation of YC-17 into methymycin and neomethymycin and narbomycin into pikromycin"  
 /note="cytochrome P450 monooxygenase"  
 /codon\_start=1  
 /transl\_table=11  
 /product="cytochrome P450 hydroxylase pikC"  
 /protein\_id="AAC6886.1"  
 /db\_xref="GI:3800840"  
 /translation="MRTQGGTASPVPVLDLQGDPAADPYPTTARLRASPAHRY  
 RPEGEVWLWGYDRARVADPRFSKDMRSTPTLTAELALNNMLSPPHTR  
 LRKLVAKEFTMRVRELLRPVQEIYDGLVADAMIAADGADMESLAMPPLPTVISEL  
 LGVPEPDRAAFVWTDAPFDPDPAQQAOTAMAAGSYLSRLIDSKRGOGEDLLALY  
 RTSDEGSRITSEELGMAHILLVAGHETTVNIAAGSYALLSPPOALALADMTLL  
 DGAVEEMLRPEGVESATIRPPEVPLDGTIYPIAGDTIVLVLAADAHRRPERPDR  
 FDIRDTAGLAFGSHGTHFCIGAPLARLRARLAVRALLERCTDLALDSFGELVWTPN  
 PMIRGLALPIRMRGRGRRGTG"  
 1505..4291  
 /gene="pikD"  
 1505..4291  
 /gene="pikD"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative transcriptional activator pikD"  
 /protein\_id="AAC6887.1"  
 /db\_xref="GI:3800841"  
 /translation="MNLVRDGEIAHLRAVLADASAGDGTLLVSGPAGSGTELLRS  
 LRRLAARETPWVSVALPQEDNDIPGLGICQLIRSRGEGATSNAYRDLIDASRRAG  
 NLTSPADAPLRVDETHRLDHWLISVSRPTPLVAVDDTLTHADTASLRFLTYCAHHRO

gene  
 CDS

GGIGFVMTERRASORAGYRVFAELLRPHCRNMWLSGLPPSGVROLLAHYGPBAER  
 RAPAATHTGNNPLLRALRODROASTTTTGAGGDEPVNDAVDLCILRSAG  
 TLETAARHATVEOSDPLVERIGTTAAAEHRIHOLAIGLDEGTLGPOPIREA  
 LODIPASERIEPILHRAAEQOLHROGADBDYVLAHNSGMAPLPLERGOALF  
 DDIRDDAFRIIEFVRSSTIDTQARLAPRILVYASRRMNHTRTALPLDSGL  
 PPSHVALIRCLVWYGRLEPAADALSRILRPSNDLLESLTRMVAALCPLES  
 PATPEPERGPVPRVLRATLQAQVFORPDNAVAQEOILOGCRLEETYEAL  
 ETALLVHADRLDRALFMSDALLAEAVERRSLGMEVFAATRAMIAIRCGDLP  
 RAEALSHAAPESWGLAVGMDLSALLACTEAGEYEOAEVLRPOVDAMFDSRGM  
 YMAHGRVYWLATGRHNAIGFPMICGSLISMNIDPSITVPWRTSAEYVLRIGNOK  
 ARALAEQOLALVPRGRSRTGRTITRVIAAVDQOARLAEAVMDLHDSGDRLEHAR  
 ALAGMSRHQOAGQNVNRYARMTARLACGMWACGAYPLAEITVFERGGRKAYSTLE  
 LPGPDVGLTSEAEERRVALAARGLTRQIARLRCLVYASTVEQHLTRYVRLVTRRA  
 DLPISLADKXSVTA"

BASE COUNT 586 a 1699 c 1485 g 572 t  
 ORIGIN

alignment\_scores:  
 Quality: 538.50 Length: 398  
 Ratio: 2.216 Gaps: 8  
 Percent Similarity: 61.055 Percent Identity: 34.673

alignment\_block:  
 US-09-724-797-36 x AF079139 ..  
 Align seg 1/1 to: AF079139 from: 1 to: 4342

15 AspyaAlArGArGsProtyrProserTyRHIStrpleuAlArGHis 31  
 188 GATTTCGGCGCGCGATCTCGACGTACGACGAGCTGTCGCGAGG 237  
 31 pProvaLHIsArG.....GlyALnHIsArGValTTrTyValS 44  
 238 TCCGGCGCCACGGGTGCGCAGCCCGAGGGGAGAGGTGCGTGG 287  
 44 eArGrpheaLAspValArGAlaValleuGlyAspGluArGrpheaLArG 60  
 288 TCGGCTACGACCGCGCGCGGTCTCGCCGATCCGCGGATCAACAAG 337  
 61 ThrGlyLeaArGrpHeTrThraSpleuValGlyProGlyLeuAl 77  
 338 GAC.....TGGCGAAGCTCCAGCAGCTCCCTGACCGAGGC 372  
 77 aglu...llevalGlyAspIleleuPheGluaspGluProAspHisG 93  
 373 CGAGCGCGCGCTCACACAAAGTCTGAGTCCGACCGCGCGGACADA 422  
 93 lyArGleuArGlyValAlGlyProAlaPheSerProserAlaLeuArG 109  
 423 CCGCGCTCGCAAGCTGTGGCCGAGTTCACCATGCGCGCGGTGAG 472  
 110 ArGleuGluProValIleAlaGlyThraValAspSpleuLeuArGProAl 126  
 473 TTGGTGGGCGCGCGGATCCAGAGATCTGACGGGCTGTGGACGCCAT 522  
 126 aleuAlaArG.....GlyAlaMetAspValValAspGluLeuAlaTyPr 141  
 523 GCGGCGCGCGCGCGAGCGCGCGGATCGATGAGGATGCTCGCTGGC 572  
 141 roLeuAlaLeuArGAlaValleuGlyLeuLeuGlyLeuProAlaAlaAsp 157  
 573 CGGTGCGGATCACCGTATCTCGAAGCTCTCGCGGTGCGCGGAG 622  
 158 TrpGlyAlaValAlGlyArGrTrpSerArGAspValGlyArGThrLeuAspAr 174  
 623 CGCGCGCGCTTCGCGCTGGACC.....GAGCG 651  
 174 gGlyAlaSerAlaGluAspMetArGrGlyHisAlaAlaIleAlaGluP 191  
 652 CTTCGTTCGCGGAGATCCCGCGAGCGCGACCGCATGCGCGAGAGA 701  
 191 heAlaAspTyValGluArGAlaLeuAlaArGArGrArGgluGly 207

```

702 TGAAGCGGCTATCTCTCCGGCTCATCGACCTCCAAAGCCGGGCGAGACGCG 751
208 GlnAspLeuAlaLeuMetLeuAspAlaHisAspArg...GlyLe 222
752 GAGAGACCTGCTAGCGCGCTCGTGGGACACGACGACGAGGAGCGGCTCCG 801
222 uMetSerArgAsnGluIleValSerThyValValThrPheIlePheThrG 239
802 GCTGACCTCCGAGGACCTCTCGGATGAGGCCACATCTCTGCTCTCCGCG 851
239 LysHisGluThrValAlaSerGlnValGlyAsnAlaValLeuSerLeu 255
852 GGACGCGAGACCGGTCATCTGATCGCCAAAGCATATGACGCTGCTC 901
256 AlaHisProAspGlnLeuAspLeuArgArgArgProAspLeuAla 272
902 TCGGACCCCGGACGAGCTGCGCCCTGGGGGCGGCACATGACGCTCTTGA 951
272 agnAlaValGluGluCysLeuArgTyrAspProSerValGlnSerAsnT 289
952 CGGCGCGGCTGAGAGACATGTTGGCTACAGAGCCCGGCTGGAAATCCGGA 1001
289 hArgArgLeuAspValAsp...ValGluLeuArgGlyArgArgLeuArg 304
1002 CTTACCGGCTTCCCGGTCCGAGCCCGTGCAGACTGCAGACGAGCATCCG 1051
305 ArgAspAspValValValValLeuAlaGlyAlaAlaAsnArgAspPro 321
1052 GCCGCGTACACGGCTCTCGTCTCGGCGCGGACGCCACGCCACCCCGGA 1101
321 gArgTyrAspArgProAspAspPheAspIleGluArgArgProValPro 338
1102 GCGGTTCCCGGACCCGACCGGCTGCGACATCCGCGGGAGACACCGCGCC 1151
338 erMetSerPheGlyAlaGlyMetArgTyrCysLeuGlySerTyrLeuAla 354
1152 ATCTGCGCTTCTGGCCACGCGATCCATCTCTGATCGAGGCGCCCTTGCC 1201
355 ArgThrGlnLeuAlaArgAlaAlaVal...AlaAlaLeuAlaArgLeuPro 370
1202 CGGTTGGAGGCGCGGATCGCGCTCGCGGCCCTTTCGGAACGCTGCCGGA 1251
370 yLeuArgLeuGlyCysAlaSerAspAlaLeuAlaTyrGlnProArgThr 387
1252 CCTGCGCCTCGAGAGCTCTCCCGGCAACCTCTGTGATCCGAACCGGA 1301
387 etPheArgGlyLeuAlaSerLeuProIleAlaPheThrProGly 401
1302 TGATTCCGGGGCTCAAGAGCCCTGCGATCCGCTGGCGGCGGAGGA 1345
seq_name: gb_Da:SEPCP450A

```

LOCUS	3078 bp	DNA	linear	BCT 26-APR-1993
DEFINITION	SERCp450A Saccharopolyspora erythraea ORF 1 gene, partial cds; cytochrome P-450 gene, complete cds; ORF 2 gene, partial cds.			
ACCESSION	M83110			
VERSION	M83110.1 GI:152682			
KEYWORDS	cytochrome P450.			
SOURCE	Saccharopolyspora erythraea (library: NRRL 2338) DNA.			
ORGANISM	Saccharopolyspora erythraea			
	Bacteria; Firmicutes; Actinobacteriae; Actinobacteridae; Actinomycetales; Pseudonocardineae; Pseudonocardaceae; Saccharopolyspora.			
REFERENCE	1 (bases 1 to 3078)			
AUTHORS	Andersen, J.F. and Hutchinson, C.R.			
TITLE	Characterization of Saccharopolyspora erythraea cytochrome P-450 genes and enzymes, including 6-deoxyerythronolide B hydroxylase			
JOURNAL	J. Bacteriol. 174, 725-735 (1992)			
MEDLINE	92121109			
FEATURES	Location/Qualifiers			
source	1..3078			

[illegible]

1152 CCAGCGAGTCAATCGAGGAAGAGCGCTGGCCGACGCGGACGAGCGGCGGGGT 1201

75 euLeuAaAGluIleValGlyAspIleIleLeuPheGlnAspGluProAsp 91

1202 TCTTCGCGCGGACCTTC...GGCCCGGACATGCTCAACACCGGACCGCGGAC 1248

92 HisGlyArgLeuArgGlyValValGlyProIaPheSerProSerIaLe 108

1249 CACACCGCGGTTCGCAACCTGATGGTCAAGCGCTTCACCGCGCGGGGT 1298

108 uArgArgLeuGluProValIleIaAGlyThrValAspAspLeuArgp 125

1299 CAGAGGCGCTCGCGCGCGGATCGAGCAAGATCAACCGAGACCTGCTCGAC 1348

125 roaIaLeuAaIaArgGlyIaMetAspValValAspGluLeuAaIaTyPro 141

1349 GCGTCGCGCGCGCGCGCTCCAAAGTGGATCGATCGACGAGTTCCGGTCCG 1398

142 LeuAaIaLeuArgAaValLeuGlyLeuLeuGlyLeuProIaIaAspTr 158

1399 CTGCCCATACGATGATCTCCGAGCTGATGGGCGCTGGAGCACTCGCGCG 1448

158 pGlyAaValAGlyArgTrpSerArgAspValAGlyArgThrLeuAspArg 175

1449 CGAGCACTTCGCGTGTGGAC...AACGCTCGGTGCAG 1486

175 LyaIaSerAaAGluAspMetArgArgGlyHisAaIaIaIaAGluPhe 191

1487 GGTGCGAGCGCGGCGGCGGCGCCAGGCGCGTGGCGGATGGTGCAG... 1533

192 AlaAspTyValGluArgAaLeuAaIaArgArgArgArgGluGlyGly 208

1534 .....TACCTGACCGAGCTGATGCGCCAAAGAGCGCACCGAGCCGGTGA 1577

208 uAspLeuLeuAaIaLeuMetLeuAaAspAa...HisAspArgGlyLeuMet 224

1578 CGACGTGCTGACCGCGCGCTGCTGAGGCGCGTGCAGAGACGGACCGCTCA 1627

224 eArgArgAsnGluIleValSerThrValValThrPheIlePheThrGlyHis 240

1628 GCGAGGCGCGAGCTGATCGCCAAATGGCTTCGCTGCTGCTGCGCGGAC 1677

241 GlnThrValAaSerGlnValGlyAsnAaValaLeuSerLeuAaIaHis 257

1578 GAGCGACGGTCAACTGATGCGCAACTGCGGCTGCTGCTGCTGGGCGA 1727

257 sProAspGlnLeuAspLeuLeuAaArgArgArgProAspLeuAaIaGln 274

1728 CCGCGACACACTCGCGCGCTGCGACAGACCGCGTGGCTGCGCGGCG 1777

274 IaValGluGluGlySLeuArgTyArgProSerValGlnSerAsnThrArg 290

1778 CGATCGAGGAGACCTCGCTACGAGAGCCCGTGGCCAAAGCGCACGTTTC 1827

291 GlnLeuAspValAspValGluLeuArgGlyArgAGlyLeuArgArgAsp 307

1828 CGGACACCGCGGAGGCG.....GTCCGGTTCGGGA 1859

307 pValIaValAa.....ValLeuAaAGlyAaIaAa 317

1860 CGTGTGATATCCCGAGGAGAGAGCTGCTGCTGGCTCGCTCGGCGGCGGA 1909

317 snArgAspProArgArgTyArgAspArgProAspAspPheAspIleAGuArg 333

1910 ACCGGGAGCGGAGCGCTTGCAGAGCACCGGCGGTGCACATCAACCGG 1959

334 AspProValProSerMetSerPheGlyAaAGlyMetArgTyArgLeuGly 350

1960 GAGACCAACGGGCGACGTGCTTCGCGCCACGCGCATCACTTCGTGCTGG 2009

350 ySerTyLeuAaArgThrGlnLeuArgAaIaValAaAa..AlaLeuA 366

2010 CGCGCGCCCTCGCGCTGGAGGCCCAAGATCGCGCTGGCGACGCTCTCG 2059

```

366 1aangLeupProGlyLeuArgLeuGlyCysAlaSerAspAlaLeuAlaTrp 382
|||||
2060 AGCGCTTCCCGGACCTCGGATGCGCGGTCCCGGACCACTCGGGTGG 2109
363 GlnPcArgThMetPheArgGlyLeuAlaSerLeuProIleAlaPheTh 399
|||||
2110 CGGTTACAGGCTCTGATGCGGGGCTCGAGAGACTGCGGCTC.....CG 2153

seq_name: gb_ba:MEPRTN8
2154 CCGCGGC 2160

seq_documentation_block:
LOCUS      MEPRTN8              342300 bp      DNA      linear      BCT 22-FEB-2001
DEFINITION Mycobacterium leprae strain TN complete genome: segment 8/10.
ACCESSION  AL583924 AL450380
VERSION    AL583924.1 GI:13093618
KEYWORDS
SOURCE
ORGANISM   Mycobacterium leprae.
            Mycobacterium leprae
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Acidimycetales; Corynebacteriaceae; Mycobacteriaceae;
            Mycobacterium.
REFERENCE  1 (bases 1 to 342300)
            Cole,S.T., Eiglmeier,K., Parkhill,J., James,K.D., Thomson,N.R.,
            Wheeler,P.R., Honore,N., Ganter,T., Churcher,C., Harris,D.,
            Mungall,K., Basham,D., Brown,D., Chillingworth,T., Connor,R.,
            Davies,R.M., Devlin,K., Duthoy,S., Feltwell,T., Fraser,A.,
            Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Lacroix,C.,
            Maclean,D., Moule,S., Murphy,L., Oliver,Qail,M.A.,
            Rajandrem,M., A., Rutherford,K.M., Rutter,S., Seeger,K., Simon,S.,
            Simmonds,M., Skelton,J., Squares,R., Squares,J.R. and Barrell,B.G.
            Taylor,K., Whitehead,S., Woodward,J.R. and Barrell,B.G.
            Massive gene decay in the leprosy bacillus
            Nature 409 (6823), 1007-1011 (2001)
            2 (bases 1 to 342300)
            Parkhill,J.
            Direct Submission
            Submitted (20-FEB-2001) Submitted on behalf of the Mycobacterium
            leprae sequencing teams, The Sanger Centre, Wellcome Trust Genome
            Campus, Hinxton, Cambridge, CB10 1SA, UK Unite de Genetique
            Molculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux
            75724, Paris Cedex, France. E-mail: parkhill@sanger.ac.uk
COMMENT
NOTES:
Details of M. leprae sequencing at the Sanger Centre are available
from http://www.sanger.ac.uk/Projects/M\_leprae/ A relational
database containing the M. leprae sequences is available from
http://genolist.pasteur.fr/Leproma/.
location/Qualifiers
    source          1..342300
                    /organism="Mycobacterium leprae"
                    /strain="TN"
                    /db_xref="taxon:1769"
                    275..430
                    /gene="ML1963"
                    275..430
                    /gene="ML1963"
                    /note="Possible pseudogene of M. tuberculosis orthologue
                    RV3463 (Best blastx score 121)"
                    /codon_start=1
                    /pseudo
                    /transl_table=1
                    /product="probable neuraminidase (pseudogene)"
                    483..1484
                    /gene="rmlB"
                    /note="ML1964"
                    483..1484
                    /gene="rmlB"
                    /note="Similar to M. tuberculosis dTDP-glucose

```

misc\_feature

4,6-dehydratase rmlB RV3464 TR:006329 (EMBL:Z95390) (331 aa); Fasta score E(): 0, 84.08 identity in 331 aa overlap, and TR:Q90556 (EMBL:U43540) (329 aa); Fasta score E(): 0, 78.2% identity in 331 aa overlap, and to others e.g. Streptococcus pneumoniae dHDP-glucose-4,6-dehydratase CPN TR:054611 (EMBL:AF030364) (349 aa); Fasta score E(): 0, 59.9% identity in 334 aa overlap. Contains Pfam match to entry PF01370 Epimerase, NAD dependent epimerase/dehydratase family.

Similar to ML0204, ML0751, ML1942 and ML2428"

/codon\_start=1  
/transl\_table=11  
/product="dHDP-glucose 4,6-dehydratase"  
/protein\_id="CAC30919.1"  
/db\_xref="GI:13093619"  
/translation="MTMRLLTVGGAGFTIGANFVNARDPDDTIVTLADMTYAGREE  
SLADYDAHRIQVGNITPTKLYEPLVAESQVYVFAEFHYNALDDEPLRFINVIG  
TPTILAVGHHSVRLAHITSTDEYGDLEDEPRTFSTPPYPPSSPATTAACDMLY  
RANVSTGYRATISCSNNIGTQYHVEREIPROTIVLTGRKPLTIGVGVKRWIMHV  
LDHNSVWQILTEKGIGRTYVLIGANGHNVLVRLTLOMDRPNDDPHATDVGIDD  
LRAIDPPTLVLYELCMAPKANFDEGLRATIDWYRNNSWMRPLDAVEAYEERG"  
495..1424

/gene="rmlB"  
/note="Pfam match to entry PF01370 Epimerase, NAD dependent epimerase/dehydratase family, score 470.30, E-value 1.6e-137"  
E-value 1.6e-137"  
1486..2094  
/gene="rmlC"  
/note="ML1965"  
1486..2094  
/gene="rmlC"  
/note="Similar to M. tuberculosis dHDP-4-dehydrorhamnose 3,5-epimerase rmlC RV3465 TR:006330 (EMBL:Z95390) (202 aa); Fasta score E(): 0, 75.4% identity in 199 aa overlap, and to many others e.g. Streptomyces glisus dHDP-4-dehydrorhamnose 3,5-epimerase SW:STRM\_STRG (P29783) (200 aa); Fasta score E(): 0, 45.6% identity in 193 aa overlap. Contains Pfam match to entry PF00908 dHDP\_sugar\_isom, dHDP-4-dehydrorhamnose 3,5-epimerase."  
/codon\_start=1  
/transl\_table=11  
/product="dHDP-4-dehydrorhamnose 3,5-epimerase"  
/protein\_id="CAC30920.1"  
/db\_xref="GI:13093620"  
/translation="MKVKEIDSGAMEITPTIHSRKNCFEEMLTASRPSAFTRHLD  
VRQANSVSACVGLHFAQVPPSOAKYVTCVTSGVFVVDVILGSPPTGCWTSITL  
LDNSNKRITITYESLHAGFLAQNSVTYMLGSAEYNPAEHAICATPDLAIDMPVAV  
DGVDTLSKRDYATNALPDLIRASGLLPTWDETQNFIANGSR"  
1492..2025

/gene="rmlC"  
/note="Pfam match to entry PF00908 dHDP\_sugar\_isom, dHDP-4-dehydrorhamnose 3,5-epimerase, score 308.80, E-value 6.5e-89"  
E-value 6.5e-89"  
2139..2624  
/gene="lpqH"  
/note="lpqH"  
2139..2624  
/gene="lpqH"  
/note="Similar to M. tuberculosis 19 kDa lipoprotein antigen precursor lpqH RV3763. Contains a possible N-terminal signal sequence and a PS00013 Proximal membrane lipoprotein lipid attachment site."  
/codon\_start=1  
/transl\_table=11  
/product="possible lipoprotein"  
/protein\_id="CAC30921.1"  
/db\_xref="GI:13093621"  
/translation="MRKHLAIAIVTIMGAGGSGGQAPTPSPSKTSTNSPTTAA  
SIPIDAAEGVIAFTBQCPYSENAKTKNGNTIQTIGASGVNTQGIHKSEFELVTC  
R"  
2169..2201  
/gene="lpqH"

misc\_feature

4,6-dehydratase rmlB RV3464 TR:006329 (EMBL:Z95390) (331 aa); Fasta score E(): 0, 84.08 identity in 331 aa overlap, and TR:Q90556 (EMBL:U43540) (329 aa); Fasta score E(): 0, 78.2% identity in 331 aa overlap, and to others e.g. Streptococcus pneumoniae dHDP-glucose-4,6-dehydratase CPN TR:054611 (EMBL:AF030364) (349 aa); Fasta score E(): 0, 59.9% identity in 334 aa overlap. Contains Pfam match to entry PF01370 Epimerase, NAD dependent epimerase/dehydratase family.

Similar to ML0204, ML0751, ML1942 and ML2428"

/codon\_start=1  
/transl\_table=11  
/product="dHDP-glucose 4,6-dehydratase"  
/protein\_id="CAC30919.1"  
/db\_xref="GI:13093619"  
/translation="MTMRLLTVGGAGFTIGANFVNARDPDDTIVTLADMTYAGREE  
SLADYDAHRIQVGNITPTKLYEPLVAESQVYVFAEFHYNALDDEPLRFINVIG  
TPTILAVGHHSVRLAHITSTDEYGDLEDEPRTFSTPPYPPSSPATTAACDMLY  
RANVSTGYRATISCSNNIGTQYHVEREIPROTIVLTGRKPLTIGVGVKRWIMHV  
LDHNSVWQILTEKGIGRTYVLIGANGHNVLVRLTLOMDRPNDDPHATDVGIDD  
LRAIDPPTLVLYELCMAPKANFDEGLRATIDWYRNNSWMRPLDAVEAYEERG"  
495..1424

/gene="rmlB"  
/note="Pfam match to entry PF01370 Epimerase, NAD dependent epimerase/dehydratase family, score 470.30, E-value 1.6e-137"  
E-value 1.6e-137"  
1486..2094  
/gene="rmlC"  
/note="ML1965"  
1486..2094  
/gene="rmlC"  
/note="Similar to M. tuberculosis dHDP-4-dehydrorhamnose 3,5-epimerase rmlC RV3465 TR:006330 (EMBL:Z95390) (202 aa); Fasta score E(): 0, 75.4% identity in 199 aa overlap, and to many others e.g. Streptomyces glisus dHDP-4-dehydrorhamnose 3,5-epimerase SW:STRM\_STRG (P29783) (200 aa); Fasta score E(): 0, 45.6% identity in 193 aa overlap. Contains Pfam match to entry PF00908 dHDP\_sugar\_isom, dHDP-4-dehydrorhamnose 3,5-epimerase."  
/codon\_start=1  
/transl\_table=11  
/product="dHDP-4-dehydrorhamnose 3,5-epimerase"  
/protein\_id="CAC30920.1"  
/db\_xref="GI:13093620"  
/translation="MKVKEIDSGAMEITPTIHSRKNCFEEMLTASRPSAFTRHLD  
VRQANSVSACVGLHFAQVPPSOAKYVTCVTSGVFVVDVILGSPPTGCWTSITL  
LDNSNKRITITYESLHAGFLAQNSVTYMLGSAEYNPAEHAICATPDLAIDMPVAV  
DGVDTLSKRDYATNALPDLIRASGLLPTWDETQNFIANGSR"  
1492..2025

/gene="rmlC"  
/note="Pfam match to entry PF00908 dHDP\_sugar\_isom, dHDP-4-dehydrorhamnose 3,5-epimerase, score 308.80, E-value 6.5e-89"  
E-value 6.5e-89"  
2139..2624  
/gene="lpqH"  
/note="lpqH"  
2139..2624  
/gene="lpqH"  
/note="Similar to M. tuberculosis 19 kDa lipoprotein antigen precursor lpqH RV3763. Contains a possible N-terminal signal sequence and a PS00013 Proximal membrane lipoprotein lipid attachment site."  
/codon\_start=1  
/transl\_table=11  
/product="possible lipoprotein"  
/protein\_id="CAC30921.1"  
/db\_xref="GI:13093621"  
/translation="MRKHLAIAIVTIMGAGGSGGQAPTPSPSKTSTNSPTTAA  
SIPIDAAEGVIAFTBQCPYSENAKTKNGNTIQTIGASGVNTQGIHKSEFELVTC  
R"  
2169..2201  
/gene="lpqH"

```

/note="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"
/complement(4768. .4995)
/gene="ML1967"
/complement(4768. .4995)
/gene="ML1967"
/note="Possible pseudogene of M. tuberculosis paralog PPE"
/codon_start=1
/pseudo
/transl_table=11
/product="PPE family protein (pseudogene)"
/complement(5464. .5649)
/gene="ML1968"
/complement(5464. .5649)
/note="Possible pseudogene of M. tuberculosis paralog PE"
/codon_start=1
/pseudo
/transl_table=11
/product="PE family protein (pseudogene)"
/complement(5661. .8043)
/note="Dispersed repeat, LEPRP, copy 6"
/complement(5848. .5946)
/gene="ML1969"
/complement(5848. .5946)
/note="Similar to Agrobacterium tumefaciens transposase TR:Q44454 (EMBL:Z18270) (366 aa) fasta scores: E(): 0.0002, 51.5% id in 33 aa, and to Pseudomonas putida transposase tpnpl TR:Q9R909 (EMBL:AF245436) (355 aa) fasta scores: E(): 0.00088, 48.5% id in 33 aa"
/codon_start=1
/pseudo
/transl_table=11
/complement(6211. .7354)
/gene="ML1970"
/complement(6211. .7354)
/note="ML1970"
/note="Similar to many e.g. Cryphonectria parasitica (Cheenut blight fungus) putative maturase TR:AF27656 (EMBL:AF218567) (778 aa) fasta scores: E(): 7.8e-11, 29.2% id in 216 aa"
/codon_start=1
/pseudo
/transl_table=11
/product="putative group II intron maturase-related protein"
/complement(6674. .6727)
/gene="ML1970"
/complement(7559. .7698)
/gene="ML1971"
/complement(7559. .7698)
/gene="ML1971"
/note="Similar to Pseudomonas putida transposase tpnpl TR:Q9R909 (EMBL:AF245436) (355 aa) fasta scores: E(): 3.4e-05, 47.8% id in 46 aa, and to Agrobacterium tumefaciens transposase TR:Q44454 (EMBL:Z18270) (366 aa) fasta scores: E(): 0.00026, 41.3% id in 46 aa"
/codon_start=1
/pseudo
/transl_table=11
/complement(8044. .8189)
/note="146 bp sequence found at 5'-end of copies 6 and 7 of LEPRP"
/complement(8461. .8691)
/gene="ML1972"
/complement(8461. .8691)
/gene="ML1972"
/note="unknown."
/codon_start=1

```

Ratio: 2.092 Gaps: 7  
Percent similarity: 62.250 Percent identity: 32.750

alignment block:

US-09-724-797-36 x MLEPRTN8/rev ..

Align seg 1/1 to reverse of: MLEPRTN8 from: 1 to: 342300

```

13 AspaLaasPvalArgArGAspProTyrProSerTyrHisTrpLeuuar 29
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134872 GACCGGGGACCGGGGCGGATCCGTTCCGTTATCGCGGTTAANTGA 134873
29 gHisAspProValHisArgGlyAlaHisArgValTrpTyrValSerArp 46
::: ||||| ::::: ||||| ::::: |||||
134822 CTACGGCGCGATGACGATACCGGGTATCCGTTGACAGTTCCTCGCGT 134773
46 heaLaasPvalArgAlaValLeuGlyAspGluArgPheAlaArgThrGly 62
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134772 TTTCTGATCGGAGAGGACATG.....CCGCATCGGTGTGCGCC 134732
63 IleArgArPheTrpThrAspLeuValGlyProGlyLeuLeuAla.... 77
||||| ::::: ||||| ::::: ||||| ::::: |||||
134731 AGCGATCGGCTCAAGCGCACATAGCGCACAGCGCATTCCTGCTGTC 134682
78 .....GluIleValGlyAspIleIleLeuPheGlnAspGluPro 91
::: ::::: ||||| ::::: ||||| ||||| |||||
134681 AGAACCGGACGCTTATGCTGCTGCTGCTTATGCTTCTGATCCACAG 134632
91 sPHISGLYArgLeuArgGlyValValGlyProAlaPheSerProSerAla 107
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134631 ATCAATACCGGCTTACGCAAGCTTGTGACAAAGCGTTCGACGAAAGTC 134582
108 LeuArgArGlyLeuGluProValIleAlaGlyThrValAspLeuLeuAr 124
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
134581 GTTCAGGCGCTGGAAGGATATGCTGCGCTGCTGATTCCTGCTGGA 134532
134 gProAlaLeuAlaArgGlyAlaMetAspValAlaAspGluLeuAlaTrp 141
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
134531 TAAAGGCTCCCGCGCTGCGCGTTCGATGCTGCTGATGCTGCTGTC 134482
141 roLeuAlaLeuArgAlaValLeuGlyLeuLeuGlyLeuProAlaAlaAsp 157
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134481 CGCTTGGCGGTGCTGTGATTTGCTGCTGCTGCTGCTGCTGCTGCTG 134432
158 TrpGlyAlaValGlyArgTrpSerArgAspValGlyArgThrLeuAsp.. 173
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134431 GCGCGAGATTGCGGTGCGGTGCTGACACTATGTCGACATGACGCC 134382
174 .....ArgGlyAlaSerAlaGluAspMetA 182
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134381 GTTCATTACCTTACCGGTACACCGCGGAGGCCACTGAAGACCGGCTGC 134332
182 rgArGlyHisAlaAlaIleAlaGluPheAlaAspTyrValGluArgAla 198
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134331 GGGCTGGT.....GTGCTGGTGGCGGATTCCTGAGACAGCTA 134294
199 LeuAlaArgArGArGArGlyGlyGlyAspLeuLeuAlaLeuMetIle 215
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
134293 GTTAAATGTCACACCGGTACACCGCGGTGAGATCTGATCCAGGCTGAT 134244
215 uAspAlaHisAspArgGly..LeuMetSerArgAspGluIleValSerT 231
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
134243 TGAGTGGACAGATCGGCTGATCAGCTGACCGAGAGAGATATATGCAA 134194
231 hrValAlaValThrPheIlePheThrGlnHisGluIleValAlaSerGlnVal 247
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134193 CGTGGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 134144
248 GlysAlaValAlaLeuSerLeuLeuAlaHisProAspGlnLeuAspLeu 264
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134143 GCGCAATGGGCTGCTGCGATGTTACGCAACCGCTGCGAAGGAAACGCT 134094
264 uArgArGlyArgProAspLeuLeuAlaGlnAlaValGluGluCysLeuArgT 281

```

```

134093 AAGTAGCAACCGCGACGCGCGCGCTGCTGTTGAGAGACACCTTCGTT 134044
281 yTrAspProSerValGlnSerAsnThrArgGlnLeuAspValGlu 297
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134043 ATGATCCAGCAATCCATTAATCGGTGCTGCTGCTGCTGCTGCTGCTG 133994
298 LeuArgGlyArgLeuArgAspAspValValValLeuAlaGlu 314
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
133993 ATCGGTGACAACTTTAACCGAGCGGTATACATGTTGTTGCTGCTGC 133944
314 yAlaAlaAsnArgAspProArgArgTyrAspArgProAspAspAsp 331
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
133943 CGCGCCAAATCGCATCCACGCTTACTACCGCCGATGAATTTGACC 133894
331 LegIuArgAspProValProSerMetSerPheGlyAlaGlyMetArgTyr 347
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
133893 CCGACCGTCCAGCTCAAGCAATTTGGCATTTGCCGCTGCACATTTT 133844
348 CysLeuGlySerTyrLeuAlaArgThrGlnLeuAlaAlaValAlaAla 364
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
133843 TGCTTAGAGCGCGCATTTGCTGCTGGAAGCAGCATGTCAATATCTGC 133794
364 Aleu...AlaArgLeuProGlyLeuArgLeuGlyCysAlaSerAspAlaL 380
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
133793 GATCAGTGGCGGCTTCCACAGGTACAGCTG.....GCCGGTGAAAT 133753
380 euaIaTyGlnProArgThrMetPheArgGlyLeuAlaSerLeuProIle 396
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
133752 TGGTCTTAAACCAAAATGTGCAATGCGGGAATGTGCGCATGCCCCGTA 133703
seq_name: gb_ba:SCE41
seq_documentation_block:
LOCUS SCE41 36028 bp DNA linear BCT 02-NOV-2000
DEFINITION Streptomyces coelicolor cosmid Ea1.
ACCESSION AL442120
VERSION AL442120.1 GI:10241774
KEYWORDS ABC transport system Arg-binding protein; ABC transport system
integral membrane protein; cytochrome P450 hydroxylase; eno;
nucleotidyltransferase; oxidoreductase; p4ae; secreted protein;
sensor kinase; transcriptional-repair coupling factor.
SOURCE Streptomyces coelicolor.
ORGANISM Streptomyces coelicolor
REFERENCE Redenbach,M., Kieser,H.M., Denapite,D., Elchner,A., Cullum,J.,
AUTHORS 1 (bases 1 to 36028)
Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 36028)
TITLE Kinsahl,H. and Hopwood,D.A.
JOURNAL A set of ordered cosmids and a detailed genetic and physical map
REFERENCE for the 8 Mb Streptomyces coelicolor A13(2) chromosome
AUTHORS Mol. Microbiol. 21 (1), 77-96 (1996)
JOURNAL 97000351
JOURNAL 2 (bases 1 to 36028)
REFERENCE Saunders,D.C. and Harris,D.
AUTHORS Unpublished
JOURNAL 3 (bases 1 to 36028)
JOURNAL Cerdano,A.M., Parkhill,J., Barrell,B.G. and Randalream,M.A.
TITLE Direct Submission
COMMENT Submitted (19-SEP-2000) Streptomyces coelicolor sequencing project,
by the BBSRC and Beowulf Genomics
details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/S.coelicolor/) CDS are
numbered using the following system eg SC7B7.01c. SC (S.
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary

```

strand). The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS. Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon. IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid E41.

## FEATURES

## SOURCE

Location/Qualifiers

1..36028

/organism="Streptomyces coelicolor"

/db\_xref="taxon:1902"

/gene="SCE41.01c"

complement(<1..951)

/gene="SCE41.01c"

/note="SCE41.01c, probable oxidoreductase (fragment), len: >317 aa; similar to SW:DHNA\_ECOLI (EMBL:V00306) Escherichia coli NADH dehydrogenase (EC 1.6.99.3) Ndb, 433 aa; fasta scores: opt: 361 z-score: 417.8 E(): 9.1e-16;

27.6% identity in 330 aa overlap and to TR:CA892372 (EMBL:AL356612) Streptomyces coelicolor putative NADH dehydrogenase SCD72A.05, 442 aa; fasta scores: opt: 1185 z-score: 1267.3 E(): 0; 57.0% identity in 321 aa overlap. Contains Pfam match to entry PF00070 pyr\_redox, Pyridine nucleotide-disulphide oxidoreductase"

/codon\_start=1

/transl\_table=11

/product="putative oxidoreductase (fragment)"

/protein\_id="CAC09533.1"

/db\_xref="GI:10241775"

/translation="MRYGEATVTVVDPKSYMTYQPLPEAAGSISPRHVVPLRPVL PKAEVLGRTVTTIDQKRVATVPLVGAPELPEDEVIVAMGAVSTPPFPGIAEGIT GMKIGESTIGLRNHVLEOLDKADSTDEETRRKALTFVVGFGAGETIGVEDMAR DAKKYNNNSREDMRITLVDAADKTLPEVGPKGIGYGEKHELEGRGEVYLVSTMSQCV DGHVVLKNGLEVDNSNTIVTAGVKNPMPALAREGLPLGPRGHVDTQATLVQOGTDYIWA AGDNAQVPDLVGVKKAENENAMPNQHALLRQAKVIGDNVI"

1..101

/note="nominal overlap with Streptomyces coelicolor cosmid SCE25"

1..36028

/organism="Streptomyces coelicolor A3(2)"

/strain="A3(2)"

/db\_xref="taxon:100226"

/clone="cosmid E41"

complement(112..798)

/gene="SCE41.01c"

/note="Pfam match to entry PF00070 pyr\_redox, Pyridine nucleotide-disulphide oxidoreductase, score 76.80, E-value

6..7e-22"

complement(1477..2418)

/gene="SCE41.02c"

complement(1477..2418)

/gene="SCE41.02c"

/note="SCE41.02c, possible hydrolase, len: 313 aa; similar to N-terminal region of SW:GPPA\_ECOLI (EMBL:M87049) Escherichia coli guanosine-5'-triphosphate, 3'-diphosphate pyrophosphatase (EC 3.6.1.40) GppA, 494 aa; fasta scores:

opt: 311 z-score: 357.2 E(): 2.2e-12; 28.4% identity in 310 aa overlap"

/codon\_start=1

/transl\_table=11

/product="putative hydrolase"

/protein\_id="CAC09534.1"

/db\_xref="GI:10241776"

/translation="MTRVAVDCGTNSIRLVADADPATGELTDDRRMTIVRLGGV DRGRILPALEKTEPFAACREAVVKAHAEIRLVATSSADANRPDVFVGLDIL GVEPEVLSGDOEAEEFTGATKEITGRALDKPYLVVDIGSGSTEEVYGEDEVRAAS VVQGVCRMEERHLVRDGAATDPPEOVAMRADIEALDLGRTVPVGEARTVIGLA GSTVTVSAIQELPEYDSAAIHHSRVSRRVETIDMLASTHAERAAVASHPERVD VIAAGSVLAIEMERGAEEVYSEHDITLDIGIAMSIA"

complement(2415..2975)

/gene="SCE41.03c"

complement(2415..2975)

/gene="SCE41.03c"

/note="SCE41.03c, conserved hypothetical protein, len: 186 aa; similar to TR:P96375 (EMBL:292539) Mycobacterium tuberculosis hypothetical 16.6 kDa protein MTC1062.24c, 135 aa; fasta scores: opt: 632 z-score: 730.7 E(): 0; 67.6% identity in 136 aa overlap"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="CAC09535.1"

/db\_xref="GI:10241777"

/translation="MQTEPTPTPTPEPDADVAFAKQOLGRPPGRLAIRHCPGQPDVETAPRLPDGCTPPTLYLVCPRASAIGILEANGVYKKEETERLADPELAAAYRA AHEDYIRRDDEIEELTGFPASAGMPDRVLCILVVAHSLAAGPVPNPLGDAIAMLPE MWKRGCVPTPTQDTEDETGEQDAQ"

complement(2425..2434)

/gene="SCE41.03c"

complement(3061..3585)

/gene="SCE41.04c"

complement(3061..3585)

/gene="SCE41.04c"

/note="SCE41.04c, hypothetical protein, len: 174 aa; similar to TR:P96376 (EMBL:292539) Mycobacterium tuberculosis hypothetical 24.6 kDa protein MTC1062.25c, 228 aa; fasta scores: opt: 273 z-score: 332.6 E(): 5.1e-11; 35.6% identity in 149 aa overlap. Contains

possible coiled-coil region at approx residues 87..106"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="CAC09536.1"

/db\_xref="GI:10241778"

/translation="MCGDVRACTGGGDDMAVKDRDRESTATRIRIIGOTARVYRSO TKQARRRRLGRALALLMVLCSLVVALAYLIRYVAGRAEIDALOREORETRQRYVD LRDKARQDDAVAEQVRLALHYVMPGRTGFVYVDPPEAEQTARAGADRPYQNV MDGVNKAQAVARQ"

complement(3639..4919)

/gene="eno"

/note="SCE41.05c"

complement(3639..4919)

/gene="eno"

/note="SCE41.05c, eno, enolase, len: 426 aa; similar to SW:ENO\_ECOLI (EMBL:X62400) Escherichia coli enolase (EC

4.2.1.11) Eno, 431 aa; fasta scores: opt: 1603 z-score: 1807.6 E(): 0; 60.6% identity in 419 aa overlap. Contains Pfam match to entry PF00113 enolase, Enol-ase and match to Prosite entry PS00164 Enolase signature"

/codon\_start=1

/transl\_table=11

/product="enolase"

/protein\_id="CAC09537.1"

/db\_xref="GI:10241779"

/translation="MPSIDVYVAREIIDSNGNPVEYVGLDGGSTGRAVPSGASNG AFEAIELRDGDPSPYLGKGVKAVLAVIYDGPETVGVDTQDQAFPDADND KGSIGANAIIIGVSAVHAASASADLPLFYRLGSPNHLIPVPMANILNGSGADSNV DIOGFMTAPIGASFSSALRMGAEVHTLTKVLKNGLATGIDDEGGFANLGSNREA IDLILEAIKRENGYVKGBOIALADVAASEFYKDGSAFECKNRSAAEMTEYVAVLVEA





Chandler, M., Choisine, N., Claudel-Renard, C., Cunnac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schlex, T., Saignes, P., Thebault, P., Whalen, M., Wincker, P., Levy, M., Weissenbach, J., and Boucher, C. A.  
Genome sequence of the plant pathogen *Ralstonia solanacearum* unpublished  
2 (bases 1 to 190050)  
Boucher, C. A.  
Direct Submission  
Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston

gene  
CDS

CDS

predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/transl\_table=11  
/evidence=not\_experimental  
/product="PROBABLE TRANSCRIPTION REGULATOR PROTEIN"  
/protein\_id="CAD17786.1"  
/db\_xref="GI:17431116"  
/translation="MKRISPPDFDMMKVKATIKRSPPAAPQGGAAISADRDGRRL  
KMLGRASISNVALNDKRSISPPRCARERKRLGCGINGVALLEIRALDGLTVIT  
GVADLRSPESPEAEKAVKQVAPCCCECHVGEEDYFLVLTNRDUNSNLAEQLL  
YLPGVQRIIRFMTLKQVLSITQLEPIA"

gene

CDS

```
complement(2336, .3508)
/ gene="RSP0647"
/ note="RS05575"
complement(2336, .3508)
```

104

**CDS**

```

complement(575, .1135)
/ gene="RSP0645"
/ note="R80557"
complement(575, .1135)
/ gene="RSP0645"
/function="miscellaneous; not classified regulator"
/ note="product confidence : probable
Gene name confidence : hypothetical
predicted by Codon usage

```

gene

CDS

```
/gene="Rsp648"  
/EC_number="4.2.1.17"  
/function="small molecule metabolism; degradation; carbon  
compounds"  
/note="Product confidence : probable  
Gene name confidence : hypothetical"
```

gene	alignment_scores	Percent Similarity	Alignment_block	gene	alignment_scores	Percent Similarity	Alignment_block
19	AspProtyrProSerTyrHisTrpLeuAlaGlnHisAspProValHisArg 35	67.442	Percent Identity: 33.850	19	AspProtyrProSerTyrHisTrpLeuAlaGlnHisAspProValHisArg 35	67.442	Percent Identity: 33.850
77427	AATCCCATATCCCTGTCAGAGACGCTGCTGCTGAGGCGCCCTTCGTGCG 77476			77427	AATCCCATATCCCTGTCAGAGACGCTGCTGCTGAGGCGCCCTTCGTGCG 77476		
35	gGlyAlaHisArgValTrpTyrValSerArgPheAlaAspValArgAlaVal 52			35	gGlyAlaHisArgValTrpTyrValSerArgPheAlaAspValArgAlaVal 52		
77477	CATGGGGCGGAATGCCGTGATGACGGGCGACATACAGCATGCTGACAGGCC 77526			77477	CATGGGGCGGAATGCCGTGATGACGGGCGACATACAGCATGCTGACAGGCC 77526		
52	AlaLeuGlyAspGluArgPheAlaArgThr.....GlyTLeuArgArg 65			52	AlaLeuGlyAspGluArgPheAlaArgThr.....GlyTLeuArgArg 65		
77527	TGCTTGACACACCGGACATGGCGAAGACCTTACATGAGAGCATGCCCTG 77576			77527	TGCTTGACACACCGGACATGGCGAAGACCTTACATGAGAGCATGCCCTG 77576		
66	PheTrpHisAspLeuValGlyProGlyLeu.....LeuAlaGluIleValAl 81			66	PheTrpHisAspLeuValGlyProGlyLeu.....LeuAlaGluIleValAl 81		
77577	CGATACGGCGAC...GAAGGCCCAACATACGCGCGTTCACAGGGGTTACG 77623			77577	CGATACGGCGAC...GAAGGCCCAACATACGCGCGTTCACAGGGGTTACG 77623		
81	YAspIleIleLeuPheGlnAspGluProAspHisGlyArgLeuArgGlyArg 98			81	YAspIleIleLeuPheGlnAspGluProAspHisGlyArgLeuArgGlyArg 98		
77624	CCGAGTGTCTCTGATGCTCAATCCGCCCATGACACAGCGCGCTGGCGGCC 77673			77624	CCGAGTGTCTCTGATGCTCAATCCGCCCATGACACAGCGCGCTGGCGGCC 77673		
98	AlValGlyProAlaPheSerProSerAlaLeuAlaArgArgLeuGluProVal 114			98	AlValGlyProAlaPheSerProSerAlaLeuAlaArgArgLeuGluProVal 114		

77674	TCATGATGCAAGTGTTCACAGCGCGGACAGATGAAATCGATGCGCAGATC	77723
115	IIeAaLgLYThValAspAspLeuAaGProAlaLeuAlaValgLYaL	131
77724	GCACAGCGCACCGCGCATGATTGATCGACGACTTCGACACAAACCGCGC	77773
131	aMetAspValValAspPglLeuAlaLysTrpProLeuAlaLeuAlaValL	148
77774	GGCGGATCTCGTGGCGGAATTCGCTCTCCCTCCCGGCGGATCATCT	77823
148	eugLYLeuLeuLeuLYLeuProAlaAlaAspTrpGLYAlaValgLYaTrp	164
77824	GCACATGATGAGCCGTCATTCGACGATGCGCATGCGCGCTGGCGTGGGG	77873
165	SerAlaGAspValGLYaTrgThrLeuAspArgGLY...AlaSerAlGLuAs	180
77874	GTCAGCAACCTCGCCAAAGGTTCATCGCGGCGCGATGAGCGCGAATGC	77923
180	pMetArgArgGLYThsAlaAlaAlaAlaGluPheAlaAspTrpAlaGLuA	197
77924	GCTGGTGGAGACGAGCGCGCGCTACGAGAACAATCCGCGCATCTTCAGA	77973
197	rgAlaLeuAlaAaArgArgArgGLYgLYgLYaLAspLeuAlaLeu	213
77974	AGGTGATTCAGAGCCCGCCCGCGCAGCCGCGCACCGCATGTTTCGATG	78023
214	MeLeuAspAlaHisAspArgGLY...LeuMetSerArgAsnGluIleVala	229
78024	CTGATCCGCGCGCGAGAAACCGCGAGACGCTGACGCGACGATGAATGCT	78073
229	IserThrValValThrPheIlePheThrGLYHisGluThrValAlaSerG	246
78074	GTCACACGTGATCGCTGCTGTATGCGCGCGCACAGACACATGCTCAACA	78123
246	IValgLYsnnAlaValLeuSerLeuAlaHisProAspPglLeuAsp	262
78124	TGATCGGCAATGCGCTGATGCGCTTGTCAATCGCAATCCGACAGCTCGAC	78173
263	LeuLeuAaArgArgArgProAspLeuAlaGlnAlaValaGLuGLuCysLe	279
78174	CTGCTCAACCGCGAGCGCGCTCGCGGATCCGACACGCGGTGCTGAATGCT	78223
279	uArgTrpAspProSerValGlnSerAsnThrArgGlnLeuAspValaSPV	296
78224	GGCTACGACGGGTGCTGTCAGATGATCCGATCCGCGCGCGCTCGAGGACG	78273
296	aIGluLeuArgGLYaTrgArgLeuAaArgArgAspAspValaValValLeu	312
78274	TCGAAGTCGAAGCGGAGGTCGTGCGCGCGGCGACAGACGGTGTCTGTATG	78323
313	AlaGLYAlaAlaAsnAaArgAspProAaArgTrpAspArgProAspAspPh	329
78324	CTCGGGCGCGCAATCGCATCGCGGCCCACTGCACCGATCGCGACCACT	78373
329	eAspIleGLuAaGAspProValProSerMetSerPheGLYAlaGlyMetA	346
78374	CGATATCGCGCGGACAGACGGCGCTGTGACAGATGCTGGCGGGGCAATCC	78423
346	rgTrpCysLeuGLYserTrpLeuAlaArgThgInLeuArgAlaAlaVal	362
78424	ATCATGAGCTCGGATGATCGCTGCGGCTGATGACATCGAAATCCGCGTG	78473
363	AlaAlaAlaLeu...AlaArgProGLYArgLeuGLYaLysAlaSerAs	378
78474	GGGCGGCTTGTGAGCGGCTCGCGAATCTGCGCGCTC...ACCAATCTCGA	78520
378	pAlaLeuAlaTrpGlnProArgThrMetPheArgGLYLeuAlaSerLeuP	395
78521	CCAGCTGACACTGGAACCAACGCGGGGCAACCTGCGGGGCGTCAACGCGTGA	78570
395	roIleAlaPhe	398







```
XX Bower SG, Perkins JB, Pero JG, Yocum RR;
XX WPI; 1995-053684/08.
XX Biotin genes, and constructs derived from Bacillus subtilis - for
PT improved production of recombinant biotin or biotin precursor for
PT use in e.g. dietary supplements
XX
XX Example II; Fig 14; 75pp; English.
XX
XX AA081792 is the B. subtilis biotin operon and flanking sequences,
CC as part of an expression vector it can be used in the recombinant
CC production of biotin (or biotin precursor protein). The biotin can
CC be used as a dietary additive in animal feeds, and as a vitamin
CC supplement for human consumption. Biotin is also useful as a
CC reagent for research, and diagnostic procedures.
XX
SQ Sequence 8478 BP; 2432 A; 1748 C; 2088 G; 2205 T; 5 other;

alignment_scores:
Quality: 609.00 Length: 396
Ratio: 2.316 Gaps: 5
Percent Similarity: 66.414 Percent Identity: 33.586

alignment_block:
US-09-724-797-36 x AA081792 ..

Align seg 1/1 to: AA081792 from: 1 to: 8478

8 AlphaheasprothrAspAlaaspValArgArgasProtyrProserTy 24
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
5493 GCATCGCACTGCATCTCTGAGTTTGGAAAAACCATATTTCTTTTA 5542
24 rHisTrpleuLeuArghisAspProvalHisArgGlyAla.....H 38
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
5543 CGACACATTTGGAGCTGTTCTATCTATTAAGGAGTTTCTTAAT 5592
38 lsaArgvalTrpTyValSerArgPheAlaaspValArgAlaValLeuGly 54
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
5593 ACCCGGGCTGTATGTCACAGATATGAAGAACGGCTGCTATTTTAAA 5642
55 AspluarGphe...AlaArgThGlylLeaArgPheTrpThrAspLe 70
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
5643 GATCGAGATTCAAAGTCCGACCCCGCTGCTGAGAGCTCAACCAATA 5692
70 uValGlyProGlyLeuLeuAlaGluIleValGlyAspIleLeuPheG 87
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
5693 TCAGGAC.....CTTTCACATGTGCAGAAATCAATGATGCTGTTTC 5733
87 lnaAspGluProAspHisGlyValArgLeuArgGlyValAlaGlyProAlaPhe 103
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
5734 AGAACCCCTGATCATAGACGATTCGCGACGCTGCCAGCGAGCGCTTT 5783
104 SerProSerAlaLeuArgArgLeuGluProValIleAlaGlyThrValas 120
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
5784 ACCCGGACGACAGACAGATTATCAGCGCATATCATTAAGACTGCCA 5833
120 pasPleuLeuArgProAlaLeuAlaArgGlyAlaMetAspValValaspG 137
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
5834 TCATTTCCTTATCAAGTCAAGTAAAAAAAAGATGAGGTCAATTTCGG 5883
137 luleuAlaTyfProLeuAlaLeuArgAlaValLeuGlyLeuLeuGlyLeu 153
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
5884 ACTTTCCTTTCTTTACAGATTGTCATACGTAACATATTAAGGCTA 5933
154 ProAlaAlaAspTrpGlyAlaValGlyArgTrpSerArgAspValGlyAr 170
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
5934 CCGGAGGAAGATAGGAGCAATTAAGAGATGGGTGCGACTGCATTCAC 5983
170 gThrLeuAspArgGlyAlaSerAlaGluAspMetArgArgGlyHisAla 187
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
5984 AACGATTGATTTTACCCGCTCAAGAAAGCATTAACAGAGGCAATATTA 6033
187 lAlleAlaGluPheAlaAspTyValGluArgAlaLeuAlaArgArgArg 203
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
6034 TGGCTGTGAGGCTATGCGATATTTCAAGAGCTGATTTCAAAAGAA 6083
204 ArgGluGlyGlylAspLeuLeuAlaLeuMetLeuAspAlaHisAspAr 220
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
6084 CCGCACCCCTCAACAGATATGATCAGCATGCTTGAAGGAGGAGAA 6133
220 gGlyLeuMetSerArgasnGluIleValSerThyValAlaThrPheIleP 237
: : : : : : : : : : : : : : : : : : : : : : : : : :
6134 GAGTAAGCTGACGAGAAAGAGCGCGCATCTACGTGCAATTCGCGGCA 6183
237 heThrGlyHisGluThrValAlaSerGlnValGlyAsnAlaValLeuSer 253
: : : : : : : : : : : : : : : : : : : : : : : : : :
6184 TCGCCGCAATGAGAACAGGCTCAATCTCATAGCAATTCATCTTTGT 6233
254 LeuLeuAlaHisProAspGlnLeuAspLeuLeuArgArgArgProAspLe 270
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
6234 CTGCTGCAGCATCCAGAACAGCTTTGAACAGTGAAGAAATCCAGATCT 6283
270 uLeuAlaGlnAlaValGluGluCysLeuArgTyArgAspProSerValGln 287
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
6284 TATTGGTACCGCAGTCAGAGAAATGTTTACGCTATGAAGCCCGCAAA 6333
287 erAspThrArgGlnLeuAspValAspValGluLeuArgGlyArgArgLeu 303
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
6334 TGAACACCCAGAGTGGCGTCAGAGATATGACATTCGCGGGTGACAGTC 6383
304 ArgArgAspAspValValAlaValLeuAlaGlyAlaAlaAsnArgAspR 320
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
6384 CGTCAAGAGAGAACAGTCTATCTTTGTTAGGAGCGGCTAATCGAGACC 6433
320 oArgArgTyArgAspArgProAspAspPheAspIleGluArgAspProValP 337
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
6434 TACATATTTACGAAACCCGATGTCTCGATATTCGAGAACTCTAATC 6483
337 roSerMetSerPheGlyAlaGlyMetArgTyCysLeuGlySerTyArgLeu 353
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
6484 CGCATCTTTCATTCGGGCATCGCATCATGTTGCTTAGGTCCTCGGTG 6533
354 AlaArgThGlnLeuArgAlaAlaVal...AlaAlaLeuAlaArgLeuPr 369
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
6534 GCACGATTGAGAACCGCAAAATGCGATTACACTCTTCGACCGAATGCC 6583
369 oGlyLeuArgLeuGlyCysAlaSerAspAlaLeuAlaTyArgInProArgT 386
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
6584 CAGCCTTAATCTT.....GCGGATTTGAATGCGCGTATCGCGCTTT 6627
386 hrMetPheArgGlyLeuAlaSerLeuProIleAlaPhe 398
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
6628 TTGGATTTTCGGGCGCTTGAAGACTCCGGTGACTTTT 6665

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/AA2000.DAT:AA75635
seq_documentation_block:
ID AAA75635 standard; DNA; 5970 BP.
XX
XX AAA75635;
XX
XX 22-JAN-2001 (first entry)
XX
XX Nucleotide sequence of ORF12 which encodes a transcriptional activator.
DE Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
KW antibiotic; C12-hydroxylase; pick; desosamine biosynthesis;
KW desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
KW picromycin biosynthesis; ss.
XX
XX Streptomyces venezuelae.
OS
XX US6117659-A.
PN
```

XX 12-SEP-2000.  
PD  
XX  
XX 27-MAY-1999; 99US-0320878.  
PF  
XX 28-MAY-1998; 98US-0087080.  
XX  
PR 22-SEP-1998; 98US-0100880.  
PR 08-FEB-1999; 99US-0119139.  
PR 20-MAY-1999; 99US-0134990.  
PR 30-APR-1997; 97US-0846247.  
PR 06-MAY-1998; 98US-0073538.  
PR 28-AUG-1998; 98US-0141908.  
XX  
XX (KOSA-) KOSAN BIOSCIENCES INC.  
PI Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;  
XX WPI: 2000-610844/58.  
DR  
XX New recombinant pick hydroxylase gene of Streptomyces venezuelae useful  
PT for converting ketolides to antibiotics and as antibiotics and  
PT intermediates in the synthesis of compounds with pharmaceutical value  
PT  
XX  
XX Disclosure; Columns 41-44; 117pp; English.  
XX  
XX The present sequence is used to produce the recombinant DNA compounds  
CC of the invention. The specification describes a recombinant DNA compound  
CC expressing recombinant polyketide synthase genes in host cells for the  
CC production of narbonolide, narbonolide derivatives and polyketides that  
CC are useful as antibiotics and as intermediates in the synthesis of  
CC compounds with pharmaceutical value. The DNA compounds may also encode  
CC a C12-hydroxylase (pick), desosamine biosynthesis and desosaminyl  
CC transferase enzymes (useful for conversion of ketolides to antibiotics),  
CC and the beta-glucosidase enzyme (involved in picomycin biosynthesis).  
CC These compounds are also useful for increasing the antibiotic activity  
CC of a compound relative to the unhydroxylated compound. The recombinant  
CC host cells are useful as genetic systems that allow rapid engineering  
CC of the narbonolide polyketide synthase. These would be valuable for  
CC creating novel ketolide analogs for pharmaceutical applications.  
XX  
XX Sequence 5970 BP; 806 A; 2219 C; 2142 G; 800 T; 3 other:  
50  
Alignment\_scores:  
Quality: 538.50 Length: 398  
Ratio: 2.216 Gaps: 8  
Percent Similarity: 61.055 Percent Identity: 34.673  
Alignment block:  
US-09-724-797-36 x AAA75635 ..  
Align seg 1/1 to: AAA75635 from: 1 to: 5970  
15 AspValArgAspProTyrProSerTyrHisTrpLeuLeuArgHisAs 31  
1422 GATTTCGGGGCGGATCCGATCCGACGTAACGAGACGTCGTCGGAGG 1471  
31 pProValHisArg.....GlyAlaHisArgValTrpTyrValS 44  
1472 TCCGGCCCAACGGGTCGACCCCGAGGAGGAGAGTGTGGCTGGCG 1521  
44 eArgpHisAlaAspValArgAlaValLeuGlyAspGlyArgpHisAlaArg 60  
1522 TCGGCTACGACCGGGCGGGGCTCTCCGCGGATCCCGGTCGACGAG 1571  
61 ThrGlyLeuArgArgpHisArgpHisArgpHisArgpHisArgpHis 77  
1572 GAC.....TGGCGCACTCCGACGACTCCCTGACCGAGGC 1606  
77 aglu...lleValGlyAspIleIleLeuPheGlnAspIleProAspHisG 93  
1607 CGAGGCCGCTCAACACACATGCTGGAGTCCGACCCCGCGGGGACACA 1656

93 lYArgLeuArgGlyValValGlyProAlaPheSerProSerAlaLeuArg 109  
1657 CCCGGCTGCGCAAGCTGTGGCGGTGAGTTCACATCGCCGCGGTGAG 1706  
110 ArgLeuGluProValIleAlaGlyThrValAspAspLeuLeuArgProAl 126  
1707 TTGCTGGCGCCCGGGGTCCAGAGATCGTCAGCGGGCTCGTGGAGCGCAT 1756  
126 AlaLeuAlaArg.....GlyAlaMetAspValValAspIleuLeuAlaTrp 141  
1757 GCTGGCGCGCCCGCCAGCGCCCGCGATGTGATGATGATGCTGCTGGCTGGC 1806  
141 TolLeuAlaLeuArgAlaValLeuGlyLeuLeuGlyLeuProAlaAlaAsp 157  
1807 CGCTGCCGATACCGGTGATCTCCGACATCTCGCGCGTGGCCGAGCCGAGC 1856  
158 TrpGlyAlaValAlaGlyArgTrpSerArgAspValGlyArgTrpLeuAsp 174  
1857 CGCGCGCTTCCGCGCTGGACC.....GAGCG 1885  
174 gGlyAlaSerAlaGluAspMetArgArgGlyHisAlaAlaIleAlaGluP 191  
1886 CTTCGTCTCCCGGAGCATCCCGCCAGCGCCAGACCGCATGGCCGAGAGA 1935  
191 heAlaAspTyrValGluArgAlaLeuAlaArgArgArgGlyGly 207  
1936 TGACCGGTATATCTCCCGGTGATCGATCCGAGCGCGGAGGAGGAGGCG 1985  
208 GluAspLeuLeuAlaLeuMetLeuAspAlaHisAspArg.....GlyLe 222  
1986 GAGACCTGCTCAGCGCGCTGTCGGAGCAGCAGCAGGAGGAGCGCTCCG 2035  
222 uMetSerArgAsnGluIleValSerThrValIleThrPheIlePheThr 239  
2036 GCTACCTCCGAGAGCTGCTCGGTATGAGCCACATCTCTGTCGCGG 2085  
239 lYHisGluThrValAlaSerGlnValGlyAsnAlaValLeuSerLeuLeu 255  
2086 GGCACGAGACACGCGTCAATGTGATCGCAAGGAGCATGTGCGGCTGCTC 2135  
256 AlaHisProAspGlnLeuAspLeuLeuArgArgProAspLeuAl 272  
2136 TCGACCCCGACCGAGCTGGCGCCCTGCGGCGGACATGACGCTTGGGA 2185  
272 aglAlaValGluGluCysLeuArgTyrAspProSerValGlnSerAsn 289  
2186 CGGCGCGGTGAGGAGATGTTGGCTAGAGGCGCGGTGGAATCCGCGA 2235  
289 hTrArgGlnLeuAspValAsp...ValGluLeuArgGlyArgArgLeuArg 304  
2236 CTTACCGCTTCCGGTGCAGCCCGTGCACCTGAGCAGCGCATATCCCG 2285  
305 ArgAspAspValValValLeuAlaGlyAlaAlaAsnArgAspProAr 321  
2286 GCGGCTACACGCGTCTGCTGTCGCGGAGCGGACCGACCGACCCCGA 2335  
321 gArgTyrArgArgProAspAspHisArgIleGluArgAspProValProS 338  
2336 GCGCTTCCCGGACCCGACCGCTTGCACATCCCGCGGAGACCGCGCGCC 2385  
338 eMetSerPheGlnAlaGlyMetArgTyrCysLeuGlySerTyrLeuAla 354  
2386 ATCTCGCTTCCGCGCAAGCATCTGTCGATCGGCGCCCGCTTGGCC 2435  
355 ArgThrGlnLeuArgAlaAlaVal...AlaAlaLeuAlaArgLeuProG 370  
2436 CGGTTGGAGGCGCGGATCGCGCTCCGCGCTTCTGGAACGCTGCCGGA 2485  
370 yLeuArgLeuGlyCysAlaSerAspAlaLeuAlaTyrGlnProArgThrM 387  
2486 CTTCGCGCTGAGAGCTCTCCCGCGGACGTCGTGTGTGATCCGAAACCGA 2535

```

387 etpheaAglyLeuAlaSerIleuProIleAlaPheThrProGly 401
|||||
2536 TGATCCGGCGGCTCAAGCCCTCGCATCCGCTCGCGGCGAGGA 2579
seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAZ56003
seq_documentation_block:
ID   AAZ56003 standard; DNA: 5970 BP.
XX
XX   AAZ56003:
XX
XX   23-MAR-2000 (first entry)
XX
XX   Contig 002 from cosmid PKOS023-27 from Streptomyces venezuelae.
XX
XX   Narbonolide polyketide synthase; PKS; cosmid PKOS023-27; contig 002;
XX   ketolide; amino transferase dehydrase; hydroxylase; plicromycin;
XX   antibiotic production; narbomycin; ds.
XX
XX   Streptomyces venezuelae.
XX
XX   Key      Location/Qualifiers
XX   CDS      complement (1..995)
XX           /*lag- a
XX           /product= PICCIY
XX           /note= "Partial amino transferase-dehydrase"
XX           /transl_except= (Pos:180..182, aa:Xaa)
XX           1356..2606
XX           /*lag- c
XX           /product= Plick
XX           /note= "Cytochrome P450 hydroxylase"
XX           2739..5525
XX           /*lag- d
XX           /product= transcripional_activator
XX           /transl_except= (Pos:4818..4820, aa:Xaa)
XX           /note= "Xaa = unknown"
XX
XX   MO9961599-A2.
XX
XX   02-DEC-1999.
XX
XX   27-MAY-1999; 99WO-US11814.
XX
XX   28-MAY-1998; 98US-0087080.
XX   28-AUG-1998; 98US-0141908.
XX   22-SEP-1998; 98US-0100880.
XX   08-FEB-1999; 99US-0119139.
XX
XX   (KOSA-) KOSAN BIOSCIENCES INC.
XX
XX   Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;
XX   WPI: 2000-072618/06.
XX   DR   P-PSDB: AAY67209, AAY67213, AAY67218.
XX
XX   New recombinant DNA encoding a domain of narbonolide polyketide
XX   synthase, for production of ketolide antibiotics -
XX
XX   Disclosure: Page 35-37, 98pp; English.
XX
XX   This is contig 002 from the recombinant cosmid PKOS023-27 DNA sequence
XX   (see AAZ56001) which contains a Streptomyces venezuelae DNA insert. The
XX   cosmid contains open reading frames which encode the various modules of
XX   the narbonolide polyketide synthase (PKS). The invention relates to
XX   recombinant DNA containing a coding sequence for a narbonolide PKS.
XX   Polyketides are compounds synthesised from 2-carbon units through a
XX   series of condensations and subsequent modifications. Modular PKSs are
XX   responsible for the production of many antibiotics including plicromycin.
XX   The narbonolide PKS consists of a loading module, six extender modules,
XX   and two thioester domains. Four proteins make up the narbonolide PKS
XX   (PICAI, PICAIi, PICAIii and PICAIv). PICAI includes the loading module
XX   and extender modules 1 and 2, PICAIi includes extender modules 3 and 4,
XX   PICAIii includes extender module 5 and PICAIv includes extender module 6

```

```

CC   and a type II thioesterase domain. The second type II thioesterase
CC   domain is found on the PICB protein. The nucleotide sequences encoding
CC   all of these proteins can be isolated in recombinant form from the
CC   recombinant cosmid PKOS023-27. Narbonolide is desosaminylated in S.
CC   venezuelae to yield narbomycin, the desosaminyl transferase enzyme is
CC   required for this conversion, and the desosaminyl biosynthetic genes are
CC   also found in cosmid PKOS023-27. The recombinant DNA of the invention is
CC   used to express, in transformed cells, narbonolide (or its derivatives)
CC   or other ketolides (particularly hybrids), which may then be converted
CC   (e.g. by other enzymes recombinantly expressed in the same hosts) to
CC   polyketide antibiotics or their intermediates. The antibiotics are useful
CC   in human or veterinary medicine.
XX
XX   Sequence 5970 BP; 806 A; 2219 C; 2142 G; 800 T; 3 other;
XX
XX   alignment_scores:
XX       Quality: 538.50      Length: 398
XX       Ratio: 2.216      Gaps: 8
XX       Percent Similarity: 61.055      Percent Identity: 34.673
XX
XX   alignment_block:
XX   US-09-724-797-36 x AAZ56003 ..
XX
XX   Align seg 1/1 to: AAZ56003 from: 1 to: 5970
XX
XX   15 AspValArgArgAspProTyrProSerTyrHisTrpLeuLeuArgHis 31
XX   ||| |||||
XX   1422 GATTTCGGCGCGCATCCGATCCGAGTACGCCAGACTCGGTGCGAGG 1471
XX
XX   31 pProValHisArg.....GlyAlaHisArgValaTrpTyrVala 44
XX   |||:|||||
XX   1472 TCCGGCCACCGGGGTGGCGACCCCGAGGGGAGAGGTGTGCTGCTG 1521
XX
XX   44 eArhPheAlaAspValArgAlaValLeuGlyAspGluArgPheAlaArg 60
XX   |||||
XX   1522 TCGGCTACGACCGGCGCGGCTCTCTCGCATCCCGCTTACACAG 1571
XX
XX   61 ThrGlyIleArgArgPheTrpThrAspLeuValGlyProGlyLeuLeuAl 77
XX   ||| ||| |||
XX   1572 GAC.....TGGCGCACTCCACGACTCCCTCGACCGAGGC 1606
XX
XX   77 aglu...IleValGlyAspIleIleLeuPheGlnAspGluProAspHisg 93
XX   |||||
XX   1607 CGAGCGCGCGCTCAACCAACAGTCGTGAGTCCGACCGCGCGGAC 1656
XX
XX   93 LysArgLeuArgGlyValValGlyProAlaPheSerProSerAlaLeuArg 109
XX   |||||
XX   1657 CCCGCTCGCCAGCTGTGGCCGTGATGATTCACATGCGCGGTCGAG 1706
XX
XX   110 ArgLeuGluProValIleAlaGlyThrValAspAspLeuLeuArgProAl 126
XX   |||||
XX   1707 TTGCTGGCGCCCGCGGTCCAGAGATGCTGACGGGTCTGTGAGCGCAT 1756
XX
XX   126 aLeuAlaArg.....GlyAlaMetAspValValAspGluLeuAlaTyrP 141
XX   |||||
XX   1757 GCTGGCGCGCGCGAGCGCGCGCGGTGATGATGATGATGATGATGATGAT 1806
XX
XX   141 roleuAlaLeuArgAlaValLeuGlyLeuLeuGlyLeuProAlaAlaAsp 157
XX   |||||
XX   1807 CGTGTCCGATCACCGTATCTCCGACTCTCGCGGTGCGCGGAC 1856
XX
XX   158 TrpGlyAlaValGlyArgTTPSerArgAspValGlyArgThrLeuAspAr 174
XX   |||||
XX   1857 CGCGCGCGCTCCCGCTGTGAC.....GACGC 1885
XX
XX   174 gGlyAlaSerAlaGluAspMetArgArgGlyHisAlaAlaIleAlaGluP 191
XX   |||||
XX   1886 CTGTGCTTCCCGGACATCCCGCGCGCGCGGACCGCATGCGCGAGA 1935
XX
XX   191 heAlaAspTyrValGluArgAlaLeuAlaArgArgArgGlyGly 207
XX   |||||
XX   1936 TGACGGGTATCTCTCCCGGTATGATGATTCACAGCGCGGACGAGCGGC 1985

```



```

208 GUAspleuLeuAlaLeuMetLeuAspAlaHisAspArg.....GlyLe 222
|||||
1986 GAGAGCTGCTGCGCTGCGGACGACGACGACGAGCGGCTCCG 2035
222 uMeSerArgansGluIleValSerThrValIValThPheIlePheThrG 239
|||||
2036 GCTAGACTCCGAGAGAGCTGCTGCTGATGCGCCACATCTCTGCTGCGG 2085
239 lYHisGluThrValAlaSerGlnValGlyAlaValAlaLeuSerLeu 255
|||||
2086 GGCACGAGACACGAGCTCAATCTGATCCCAACGACATGATACGCTCTC 2135
256 AlaHisProAspGlnLeuAspLeuLeuArgArgArgProAspLeuAla 272
|||||
2136 TCGACACCCCGACGACGCTGCGCCGCGCCACATGACGCTCTTGGGA 2185
272 aGlaAlaValaGluGluCysLeuArgTyrAspProSerValGlnSerAsn 289
|||||
2186 CGGCGCGGTGAGAGAGATGTTGCCCTACGAGGCGCCGCTGGAATCCGCG 2235
289 hrArgGlnLeuAspValAsp...ValGluLeuArgGlyArgArgLeuArg 304
|||||
2236 CCFACCGCTTCCGCTGCGACGCTGACCTGACGACGACGACGCTCATCCG 2285
305 ArgAspAspValValValValLeuAlaGlyAlaAlaAsnArgAspProAr 321
|||||
2286 GCCGGTGACAGGCTGCTGCTGCTGCGCCGACGCGCCACCGACCCCGGA 2335
321 gArgTyrAspArgProAspAspPheAspIleGluArgAspProValProS 338
|||||
2336 GCGCTTCCCGGACCCGCGCCGCTTGCATCGCGCGGACCGCCGCGGCC 2385
338 erMeSerPheGlyAlaGlyMetArgTyrCysLeuGlySerTyrLeuAla 354
|||||
2386 ATCTGCGCTTCCGCGACGACGATCACTTCTGCATCGCGCGCCCTTGCGC 2435
355 ArgThrGlnLeuArgAlaAlaVal...AlaAlaLeuAlaArgLeuProG 370
|||||
2436 CGGTTGGAGGCCCGGATCGCGCTGCGCCCTTCTCGACACGCTCCGCGGA 2485
370 yLeuArgLeuGlyCysAlaSerAspAlaLeuAlaTyrGlnProArgThr 387
|||||
2486 CCTGCGCTTGGAGCGTCTCCCGGCGGACGATGCTGTGATCGAACC CGA 2535
387 etPheArgGlyLeuAlaSerLeuProIleAlaPheThrProGly 401
|||||
2536 TGATCCGCGGCTCAAGGCCCTGCGCATCGCTGCGCGCGCGAGGA 2579
seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DMT:AAZ87301
seq_documentation_block:
ID AAZ87301 standard; DNA; 1251 BP.
XX
AC AAZ87301;
XX
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae macrolide biosynthetic gene pikC, SEQ ID NO:38.
XX
KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW neomethymycin; narbomycin; polyhydroxalkanoate monomer synthase;
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma,
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolemia; crop protection agent; ds.
XX
OS Streptomyces venezuelae ATCC15439.
XX
FH Key location/Qualifiers
FT CDS 1..1251
FT /*Lag= a
FT /product= "pikC"
FT /transl_except= (pos:307..309, aa:Leu)

```

```

XX MO200000620-A2.
XX
XX 06-JAN-2000.
XX
XX 25-JUN-1999; 99WO-US14398.
XX
XX 26-JUN-1998; 98US-0105537.
XX
XX (MINU) UNIV MINNESOTA.
XX
XX Sherman DH, Liu H, Xue Y, Zhao L;
XX
XX WPI: 2000-160679/14.
XX
XX P-PSDB: AAY77196.
XX
XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
XX synthesis of methymycin and pikromycin -
XX
XX Claim 15; Page 428-429; 438pp; English.
XX

```

The invention relates to an isolated and purified nucleic acid segment comprising a desosamine biosynthetic gene cluster, a fragment or its biologically active variant, where the nucleic acid sequence is not derived from the *eryC* gene cluster of *Saccharopolyspora erythraea* or Streptomyces antibiotics. The invention also relates to a macrolide biosynthetic gene cluster, or fragments thereof. The macrolide biosynthetic gene cluster encodes proteins which synthesise methymycin, pikromycin, neomethymycin, narbomycin or a combination of these compounds. Recombinant or augmented cells comprising the desosamine and/or macrolide biosynthetic gene clusters are useful for the production of biologically active macrolides. The macrolide biosynthetic proteins are useful for synthesis of methymycin, pikromycin, neomethymycin and narbomycin. The alternative termination of polyketide synthesis may be useful to prepare novel antibiotics and polihydroxyalkanoate (PHA) monomers. The compounds produced by the recombinant host cells are useful as biopolymers, e.g., in packaging or biomedical applications, to engineer PHA monomer synthases or to prepare biologically active agents, such as chemotherapeutics, immunosuppressants, agents to treat asthma, chronic obstructive pulmonary disease as well as other diseases involving respiratory inflammation, cholesterol-lowering agents or macrolide-based antibiotics which are active against a variety of organisms, e.g., bacteria, including multi-drug resistant pneumococci and other respiratory pathogens, as well as viral parasitic pathogens, or as crop protection agents (e.g., fungicides or insecticides) via expression of polyketides in plants. Sequences AAZ87295-287302 represent macrolide biosynthetic genes from Streptomyces venezuelae ATCC 15439, which encode proteins AAY77190-Y77197.

Sequence 1251 BP; 171 A; 484 C; 413 G; 183 T; 0 other;

alignment\_scores:  
Quality: 531.50 Length: 398  
Ratio: 2.196 Gaps: 8  
Percent Similarity: 60.804 Percent Identity: 34.422

alignment\_block:  
US-09-724-797-36 x AAZ87301 ..

Align seg 1/1 to: AAZ87301 from: 1 to: 1251

```

15 AspValArgArgAspProTyrProSerTyrHisTyrLeuLeuArgHisAs 31
|||||
67 GATTCCGCGCGCGATCCGATCCGATCCGACGACGACGCTCCGAGGG 116
31 pProValHisArg.....GlyAlaHisArgValTyrTyrValS 44
|||||
117 TCGCGCCACGCGGTGCGACCCCGGAGGGGAGAGATGTGCTGTGCTG 166
44 erArgPheAlaAspValArgAlaValaLeuGlyAspGluArgPheAlaArg 60
:::

```

```

167 TCGGCTACGACCGGGCGGGCGGTCTCGCCGATCCCGGGTTCAGCAAG 216
61 ThrGlyIleArgArgPheTrpThrAspLeuValGlyProGlyLeuAla 77
217 GAC.....TGGCGCAACTCCACGACTCCCTGACCGAGGC 251
77 aglu...lIeValGlyAspIleIleLeuPheGlnAspGluProAspHisG 93
252 CGAGGCCCGCGCTCAACACACATGCTGGAGTCCGACCGCGCGGCGACA 301
93 lYArgLeuArgGlyValValGlyProAlaPheSerProSerAlaLeuArg 109
302 CCGCGCCCGCGCAAGCTGGTGGCGGTGATCCACATGCGCGCGGTGAG 351
110 ArgLeuGluProValIleAlaGlyThrValAspAspLeuLeuArgProAl 126
352 TTGGTGGCGGGCGCGGGTCCAGAGATGCTGCACGGGCTGGAGCGCCAT 401
126 AlEuaAlaArg.....GlyAlaMetAspValValAspGluLeuAlaTyrP 141
402 GCTGGCGCGCGCCGACGCGCGCGCGATGATGAGTCCCTGGCGTGGC 451
141 rOlEuaAlaLeuArgAlaValLeuGlyLeuLeuGlyLeuProAlaAlaAsp 157
452 CGCTGCCGATCACCGTGATCTCCGAACTCTCGGCGGTGCCGACCGGAC 501
158 TrpGlyAlaValGlyArgTrpSerArgAspValGlyArgThrLeuAspAr 174
502 CGCCCGCGCTTCGCGTCTGGACC.....GACGC 530
174 gGlyAlaSerAlaGluAspMetArgArgGlyHisAlaAlaIleAlaGluP 191
531 CTTCGTCTCCGCGACGATCCCGCGCGCGCGACCGCATGCGCGCGAGA 580
191 heAlaAspTyrValGluArgAlaLeuAlaArgArgArgGluGly 207
581 TGACGGGCTATCTCTCCCGCTCATCGATCCCAAGCGCGGCGAGAGCG 630
208 GluAspLeuAlaLeuMetLeuAspAlaHisAspArg.....GlyLe 222
631 GAGGAACTGCTCAAGCGCGCTCGTCCGACCAAGGAGAGAGAGAGCGCTCCG 680
222 uMeSerArgAsnGluIleValSerThrValValThrPheIlePheThrG 239
681 GCTGACCTCCGAGAGCTGCTCGGTANGCCCAATCTCTGCTGCGCG 730
239 lYHisGluThrValAlaSerGluValGlyValAlaValLeuSerLeu 255
731 GGCACGAGACACACGTCATCTGATGCCCAACGCGATGATGCGGTGCTC 780
256 AlAHisProAspGluLeuAspLeuLeuArgArgProAspLeuAla 272
781 TCCGACCCCGACAGCTGCGCGCGCGCGCGACATGACGCTCTTGG 830
272 aglAlaValGluGluCysLeuArgTyrAspProSerValGlnSerAsnT 289
831 CGGCGCGGTGAGAGATGTTGGCTACGAGGCGCGGTGGAATCCGCGA 880
289 hrArgGlnLeuAspValAsp...ValGluLeuArgGlyArgLeuArg 304
881 CTrACCGCTTCGCGGTGAGACCGCTGACTGGACGCGACGCGATCCCG 930
305 ArgAspAspValValValValLeuAlaGlyAlaAlaAsnArgAspProAr 321
931 GCGCGTACACAGGTCTGCTGCTGCGGAGCGCGCCACCGCACCCCGCA 980
321 gArgTyrAspArgProAspAspPheAspIleGluArgAspProValProS 338
981 GCGCTTCCCGACCGCACCGCTTCGACATCCCGCGGACACCGCGCGCC 1030
338 erMetSerPheGlyAlaGlyMetArgTyrCysLeuGlySerIleAla 354
1031 ATCTCCGCTTCGCGCACGCGATCTCTGCACTGGCGCGCGCGCTTGGCC 1080

```

```

355 ArgThrGlnLeuArgAlaAlaVal...AlaAlaLeuAlaArgLeuProG 370
1081 CGGTTGGAGGCCCGGATCGCGCTCCGCGCTTCTCGAACGCTGCCCGA 1130
370 yLeuArgLeuGlyCysAlaSerAspAlaLeuAlaTyrGlnProArgThiM 387
1131 CCTCGGCCCTTGACCTCTCCCGCGCGCAACTCGTGTGTTCGACACCGCA 1180
387 eTrPheArgGlyLeuAlaSerLeuProIleAlaPheThrProGly 401
1181 TGATTCGCGGCTCAAGCGCCCTCGATCCGCTGGCGGCGGAGGA 1224

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.AAC5790
seq_documentation_block:
ID AAC5790 standard; DNA; 1233 BP.
XX
AC AAC5790;
XX
DT 19-JAN-2001 (first entry)
XX
DE Hydroxylase orf3 involved in MC biosynthesis.
XX
KW Mitomycin; biosynthesis; mitosome ring system; antibiotic; anti-cancer;
KW anti-inflammatory; immune-enhancer; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW fungicide; pesticide; ds.
XX
OS Streptomyces lavendulae.
XX
PN WO200053737-A2.
XX
PD 14-SEP-2000.
XX
PF 10-MAR-2000; 2000WO-US06394.
XX
PR 12-MAR-1999; 99US-026965.
XX
PA (MNU) UNIV MINNESOTA.
PA (SHER) SHERMAN D H.
PA (MAOY) MAO Y.
PA (VARO) VAROGLU M.
PA (HEMM) HE M.
PA (SHEL) SHELDON P C.
XX
PI Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;
XX
DR WPI; 2000-601980/57.
XX
PT Novel nucleic acid molecule comprising mitomycin biosynthetic gene
PT cluster useful for cloning mitomycin biosynthetic genes for elucidating
PT the molecular basis of mitosome ring system biosynthesis
XX
PS Example 1; Page 255; 399pp; English.
XX
CC This invention relates to isolated and purified nucleic acid molecules
CC from the mitomycin biosynthetic gene cluster. Mitomycins are a group of
CC natural products that contain a variety of functional groups, including
CC amino benzquinone and axiridine ring systems. The S. lavendulae
CC mitomycin biosynthetic gene cluster comprises 47 mitomycin genes
CC spanning 55kb of DNA. The invention includes an expression cassette
CC comprising a mitomycin biosynthetic gene operably linked to a promoter,
CC and host cells transformed with the cassette. The nucleotide, and protein
CC sequences and the transformed host cells of the invention result in
CC antitumor, anti-inflammatory, cytostatic, immunomodulatory, and
CC antibiotic activities. The nucleotide sequences are used to elucidate the
CC molecular basis for the biosynthesis of the mitosome ring system, as well
CC as to engineer the biosynthesis of novel natural products, e.g.
CC antibiotics, anti-inflammatory agents, anti-cancer agents,
CC immune-enhancers, immunosuppressants, agents to treat asthma, chronic
CC obstructive pulmonary disease as well as other disease involving
CC respiratory inflammation, or cholesterol-lowering agents or as crop

```



```

FT      /note= "Encodes a protein containing 268 amino acids
FT      significantly similar to known S-adenosyl
FT      methionine-dependent methyl transferase"
FT      CDS
FT      2593..4011
FT      /tag= c
FT      /label= Gene_3
FT      /note= "Encodes a protein containing 472 amino acids"
FT      CDS
FT      4013..4999
FT      /tag= d
FT      /label= Gene_4
FT      /note= "Encodes a protein containing 328 amino acids"
FT      CDS
FT      5071..6085
FT      /tag= e
FT      /label= Gene_5_(part)
FT      /note= "Partial sequence for gene 5; full length
FT      protein contains 366 amino acids and is
FT      significantly similar to amino transferase
FT      enzymes, such as the Dnr J protein"
FT
PN      W09708323-A1.
XX      06-MAR-1997.
XX      19-AUG-1996; 96WO-EP03643.
XX      30-AUG-1995; 95EP-0810534.
XX      (CIBA ) CIBA GEIGY AG.
XX
PI      Bietenhader J, Engel N, Pospiech A, Schupp T, Toupet C;
DR      WPI, 1997-179280/16.
XX
PT      Indole-carbazole alkaloid biosynthesis gene cluster - especially
PT      coding for the antibiotic staurosporin from Streptomyces
PT      longisporoflavus
XX
PS      Claim 7; Page 37-41; 55pp; English.
XX
XX      The present sequence represents the 6.5kb PvuII fragment of Streptomyces
XX      longisporoflavus R19 which is involved in the biosynthesis of
XX      indole-carbazole alkaloids (ICA). The sequence contains five functional
XX      fragments as indicated in the features table. The DNA or a hybrid
XX      vector containing it can be used to prepare an ICA or derivatives and
XX      precursors, either by allowing production in previously incapable
XX      organisms or by improving yields. In particular, the antibiotic
XX      staurosporin can be produced. Staurosporin is known to have inhibitory
XX      activity against fungi, yeasts, and Ca2+/phospholipid-dependent
XX      serine/threonine protein kinases (PKCs). Staurosporin also has
XX      antiproliferative activity and can inhibit platelet aggregation.
XX      The present sequence can also inactivate ICA biosynthesis genes and
XX      can be used in PCR amplification. An advantage of this is that
XX      productivity of staurosporin-synthesising Streptomyces is improved
XX      over natural strains yielding only low concentrations.
XX
XX      Sequence 6085 BP; 882 A; 2374 C; 1992 G; 833 T; 4 other;
XX
XX
XX      Alignment_scores:
XX      Quality: 508.50      Length: 407
XX      Ratio: 2.101      Gaps: 11
XX      Percent Similarity: 59.459      Percent Identity: 36.364
XX
XX      alignment_block:
XX      US-09-724-797-36 x AAT70153 ..
XX
XX      Align seg 1/1 to: AAT70153 from: 1 to: 6085
XX
XX      20 ProTyrProSerTyrHisTrpLeuAlaArgHisAspProValHisArg.. 35
XX      ||||| ||||| :: |||||
XX      441 CCGTACCGGCTACCGGCGGAGCGCCCGCGGTCAATCGCAC 490
XX      36 .....GlyAlaHisArgValTrpTyrValSerArgPheAla 48

```

```

491 GCGCGCGCGCCCGGAAAGCCTGACACCTACTAGTGTTCACCTACGACG 540
FT      ||| ..... |||
FT      48 spValArgAlaValLeuGlyAspGluArgPheAlaArgTrgIleArg 64
FT      ||||| ||||| ..... |||||
FT      541 ACGTGTCTCCGCTCTGCTCCACCGGCGGTTCGCGCCGACAGCC..... 584
FT      ||| ..... |||
FT      65 ArgThrTrpThrAspLeuValGlyProGly..... 74
FT      ||| ..... |||
FT      585 CGCGTGGCTCTCGCGGACACCGCGCCGACACCGCGCCGCTCCGATGCC 634
FT      ||| ..... |||
FT      75 .....LeuAlaGluIleValGlyAspIleIleuPheG 87
FT      ||| ..... |||
FT      635 GCGCGAGCAGCGCCGCTCGCGGACCGTGTGACAGAACTGCTGTCCTCC 684
FT      ||| ..... |||
FT      87 LnsAspGluProAspHisGlyArgLeuArgGlyValValGlyProAlaPhe 103
FT      ||| ..... |||
FT      685 TCGACCCCGCGCGGACACCGCACTGCGCTCCCTGACCGCGGACTTC 734
FT      ||| ..... |||
FT      104 SerProSerAlaLeuArgArgGluIleProValIleAlaGlyThrValAs 120
FT      ||||| ||||| ..... |||||
FT      735 TCACCTCTGATCTGTCACCGGCTGCGCCCGCATGCCAGACCGGAGG 784
FT      ||| ..... |||
FT      120 PASpLeuLeuArgProAlaLeuAlaArgGlyAlaMetAspValValAspG 137
FT      ||| ..... |||
FT      785 CGAACTCTGGACCGGCTCGGACACACCGCGCGCGGATCTCTCGAGG 834
FT      ||| ..... |||
FT      137 LuLeuAlaTyrProLeuAlaLeuArgAlaValLeuGlyIleuGlyLeu 153
FT      ||| ..... |||
FT      835 GTTCGGCGCGCCCTCCCGGATCTCTGTCATCTCCGACTCTGCGGCATC 884
FT      ||| ..... |||
FT      154 .ProAlaAlaAsp.....TrpGlyAlaValGlyArgTrpSer..ArgAs 167
FT      ||| ..... |||
FT      885 CCCCCCGGAGAGACCAACCTGTGTGGCGCCCAAGCGGTGGCCCTTCAGGA 934
FT      ||| ..... |||
FT      167 pValGlyArgThrLeuAspArgGlyAlaSerAlaGluAspMetArgArg 184
FT      ||| ..... |||
FT      935 GCGCGGCGACACAGCTC.....GCGCGGCGCGGACGATACGACGCGG 975
FT      ||| ..... |||
FT      184 LysHisAlaAlaIleAlaGluPheAlaAspTyrValAlaArgAlaLeuAla 200
FT      ||||| ||||| ..... |||||
FT      976 CCGAGCGCGGCTCCAGAGATTCACCGCTACTTCCGCGGAGAGGTGAC 1025
FT      ||| ..... |||
FT      201 ArgArgArgArgGluGlyGlyGluAspLeuLeuAlaLeuMetLeuAspAl 217
FT      ||||| ||||| ..... |||||
FT      1026 CGCGCGCGCGGCGGACCGCGCATGCTCTCCCTCTCCGCGGC 1075
FT      ||| ..... |||
FT      217 AHisAspArgGlyLeu...MetSerArgAsnGluIleValSerThrValY 233
FT      ||||| ||||| ..... |||||
FT      1076 CCGGAGACCGGATACCGCTAGCGTGGACGCGCATCTCGCACCTCGG 1125
FT      ||| ..... |||
FT      233 AlThrPheIlePheThrGlyHisGluThrValAlaSerGlnValGlyAsn 249
FT      ||| ..... |||
FT      1126 TCATCTGTCTACCGCGCGGCGGACAGACACCACTGCTCTCCGACG 1175
FT      ||| ..... |||
FT      250 AlaValLeuSerLeuLeuAlaHisProAspGlnLeuAspLeuLeuArg 266
FT      ||||| ||||| ..... |||||
FT      1176 GCGGTCTCTCCCTCGCGCCACCTGACTGCTCTGAGAGACTCGGCAC 1225
FT      ||| ..... |||
FT      266 GArgProAspLeuLeuAlaGlnAlaValGluGluCysLeuArgTyrAsp 283
FT      ||||| ||||| ..... |||||
FT      1226 CACACCGGAGTGCACACCGCGCGGTGCGAAGAGTGAAGCGGTACGACC 1275
FT      ||| ..... |||
FT      283 roSerValGlnSerAsnThrArgGlnLeuAspValAspValGluLeuArg 299
FT      ||| ..... |||
FT      1276 CGCCGCTGAGCGGTGAGCGCTGCGCTGACAGACATCCGCGTCCGCC 1325
FT      ||| ..... |||
FT      300 GlyArgArgLeuArgArgAspValValValValLeuAlaGlyAlaAl 316
FT      ||| ..... |||
FT      1326 GACCAAGACATCCCGCGGCGGACCGCGGTGTGCGCTCTCGGCTCGGC 1375
FT      ||| ..... |||
FT      316 aAsnArgAspProArgArgTyrAspArgProAspAspPheAspIleGluA 333
FT      ||||| ||||| ..... |||||

```







alignment\_scores:                      Length:                      382  
     Quality:                      498.50  
     Ratio:                        2.060  
     Percent Similarity:        63.351                      Percent Identity:        34.555

alignment\_block:

US-09-724-797-36 x AAF81357        ..

Align seq 1/1 to: AAF81357        from: 1 to: 1257

```

32  ProValHisArgGlyAla...HisArgValTrpTyrValSerArgPheal 47
   |||:|||||:||||| ||| ||||| |||:|||||:
91  CCGCGCATGCGCGCGCGAGCATCCGGTGTGGGTGACGCGCATACG 140
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
47  aAspValArgAlaValLeuGly.....AspGluArgP 58
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
141 CGAGCGCGCGAAGGTCTCAACCATCCGGCGCTCCGCGCGAGCGCGG 190
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
58  heAla.....ArgThrGlyLeuArgArgPheTrpThr 68
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
191 AGGCGCGCAACTCTACCGCAAGGTACCGGACCGCGCGCGG... 237
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
69  AspleuValGlyProGlyLeuLeuAlaGluIleValGlyAspIleIle 85
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
238 .....ATCGCGGAGGACTC.....AGCCACACATGCT 266
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
85  uPheGluAspGluProAspHisGlyArgLeuArgGlyValGlyProA 102
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
267 CACCTCGACCGCGCGCATACCGCTCGCTGCTGCTGGCGCG 316
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
102 LaphSerProSerAlaLeuArgArgLeuGluProValIleAlaGlyThr 118
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
317 CGTTACACCGCGCGAGGTGAGCGCTGCACCGCATATAGAACGATC 366
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
119 ValAspAspleuLeuArgProAlaLeuAlaGlyAlaMetAspVala 135
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
367 ACCGAGGATGCTGTGACGCCATGCGCGCGCACAGCGCATGAT 416
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
135 LAspGluLeuAlaTrpProLeuAlaLeuArgAlaValLeuGlyLeuLeu 152
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
417 GCGCGACTTCGCGATCCCGCTGACCATCGCGGTATCTTGAGCTGCTGG 466
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
152 LyleuProAlaAlaAspTrpGlyAlaValGlyArgTrpSerArgAspVal 168
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
467 GCATTCCGAGCGCGAG.....CGCGACAC 492
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
169 GlyArg..ThrLeuAspArgGlyAlaSerAlaGluAspMetArgArgG 184
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
493 GCGCGCCAGTCTGGAGCGCGCAGCGAATGCTGTGCGCGGAGGAGGC 542
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
184 yHAlaAlaIleAlaGluPheAlaAspTyrValGluArgAlaLeuAla 201
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
543 CCAGGCCCTGGCCGATGCGCAGGTGCTGCTGCTGCTGCTGCTGAGG 592
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
201 rGArgArgArgGluGlyGlyLeuAspLeuAlaLeuMetLeuAspAla 217
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
593 CCAAGCGCGCGAGCGCGCGACGCTACAGCGCGGCTGCTGCTGAGGCC 642
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
218 HisAspArg...GlyLeuMetSerArgAsnGluIleValSerThrVala 233
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
643 GCCGCGAGAGCGCGCATGTGAGAACCGGAACTGCTTCATGAGGCCA 692
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
233 LThrPheIlePheThrGlyHisGlyThrValAlaSerGluValGlyAsn 250
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
693 CCGTGATGATGAGCGGCTTCGAGACCATGATGATGATGATGATGAT 742
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
250 LavalLeuSerLeuLeuAlaHisProAspGluLeuAspLeuLeuArgArg 266
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
743 CGCTGTGTACCTGTGTGTCAACCGGAGCAACTGGCTTCTGCGGGCG 792
   |||:||||| |||:|||||: |||:|||||: |||:|||||:
267 ArgProAspLeuLeuAlaGluAlaValGluGlyCysLeuArgTyrAspPr 283
   |||:||||| |||:|||||: |||:|||||: |||:|||||:

```

```

793 CAGCGGAACCTCTGCCACAGCCATGGAGAACTGTCCGCCACGACAG 842
283 oSerValGlnSerAsnThrArgGlnLeuAspVal...AspValGluLeuA 299
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
843 TCCGCTGCGCGCTGATGTGCTTCACCGTGAAGACATGGAGACTGG 892
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
299 rGlyArgArgLeuArgArgAspAspValValValLeuAlaGlyAla 315
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
893 ACGGGGTACCATTTCCCGCGCGGAGAAATACCTGCTGTCCACCTGAC 942
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
316 AlaAsnArgAspProArgArgTyrAspArgProAspAspPheAspIle 332
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
943 GCCAACCCAGCCGAGCGCTTCGACGATCCCGACCCCTGACCTTAC 992
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
332 uArgAspProValProSerMetSerPheGlyAlaGlyMetArgTyrCysL 349
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
993 CCGCAACACCGATGGCATCTTGCTTACGGCTTCCGCTGACCTTCTCG 1042
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
349 euGlySerTyrLeuAlaArgThrGlnLeuArgAlaAlaVal...AlaAla 364
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
1043 TCGGCGCTCGCTGGCGCGGTGAGAGGGGCGATCGCATCCAGCGCCTG 1092
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
365 LeuAlaArgLeuProGlyLeuArgLeuGlyCysAlaSerAspAlaLeuAl 381
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
1093 CTCGCGCGCTTCCCGACCTCCAGTTGGCGTGCCCGCACGCGGACTCA 1142
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
381 aTyrGlnProArgThrMetPheArgGlyLeuAlaSerLeuProIle 396
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
1143 GTGCTGCGGATCACCTTCTCTCGCGCTGTGATCAGCGCGGTG 1188
   |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAA09469
seq_documentation_block:
ID AAA09469 standard; DNA: 50937 BP.
xx
AC AAA09469;
xx
DT 29-AUG-2000 (first entry)
xx
DE Streptococcus oleandomycin gene cluster.
xx
KW oleandomycin; oleandomide; polyketide synthase; oleA; oleAII; oleAIII;
KW PKS; type I; 8,8a-deoxyoleandomide synthase; modular; ketosynthase;
KW acyl-transferase; acyl carrier protein; inactivated; polyketide;
KW macro lactone; antibiotic; motilide; erythromycin; ss.
xx
OS Streptococcus antibioticus.
xx
FH key
FT 152..1426
   Location/Qualifiers
   /tag= a
   /label= oleI
   complement (1528..2637)
   /tag= b
   /label= olen2
   complement (2658..4967)
   /tag= c
   /label= oler
   5772..18224
   /tag= d
   /label= ORF1
   /product= 8,8a-deoxyoleandomide_synthase_1
   18267..29717
   /tag= e
   /label= ORF2
   /product= 8,8a-deoxyoleandomide_synthase_2
   29787..40346
   /tag= f
   /label= ORF3
   /product= 8,8a-deoxyoleandomide_synthase_3
   40625..41830
   /tag= g
   /label= oleP1
FT

```



FT CDS 41878..43158  
FT /tag= h  
FT /label= olegl  
FT 43163..44443  
FT /tag= i  
FT /label= oleG2  
FT 44433..45173  
FT /tag= j  
FT /label= oleM1  
FT 45251..46411  
FT /tag= k  
FT /label= oleY  
FT 46491..47714  
FT /tag= l  
FT /label= oleP  
FT CDS complement (47808..49517)  
FT /tag= m  
FT /label= oleB  
XX WO200026349-A2.  
XX 11-MAY-2000.  
XX 22-OCT-1999; 99WO-US24478.  
XX 29-OCT-1998; 98US-0106100.  
XX 16-FEB-1999; 99US-0120254.  
XX (KOSA-) KOSAN BIOSCIENCES INC.  
XX Betlach MC, Shah SK, McDaniel R, Tang L;  
XX WPI: 2000-365602/31.  
XX P-PSDB: AAY92707, AAY92708, AAY92709.  
XX Recombinant DNA compound encoding oleandolide polyketide synthase for  
XX synthesizing polyketides comprising a coding sequence for a domain of a  
XX loading module or any one of extender modules  
XX Disclosure; Page 14-26; 86pp; English.  
XX This is part of the Streptococcus antibioticus oleandomycin gene cluster.  
XX The oleandolide polyketide synthase (PKS), also known as  
XX 8,8a-deoxyoleandolide synthase, is encoded by three open reading frames  
XX (ORF), designated oleA1, oleAII and oleAIII. The PKS is a type I  
XX "modular" enzyme, where each ORF encodes 2 extender modules and  
XX of at least a ketosynthase (KS). Each module is composed  
XX of an acyl carrier protein (ACP) domain. The oleandolide PKS loading module  
XX contains an inactivated KS, called KS-0, where 0 is the abbreviation for  
XX glutamine. The large multifunctional PKS enzymes catalyze the biosynthesis  
XX of polyketide macrolactones through multistep pathways involving  
XX decarboxylative condensations between acylthioesters followed by cycles  
XX of varying beta-carbon processing activities. The macrolide product of  
XX the PKS, 8,8a-deoxyoleandolide, is further modified by epoxidation and  
XX glycosylation to yield oleandomycin, an antibacterial polyketide. The  
XX invention concerns an isolated recombinant DNA compound, comprising a  
XX coding sequence for a domain of loading module or any one of extender  
XX modules 1-4 or 1-6, including an oleandolide PKS operably linked to a  
XX promoter. Also discussed are recombinant oleandolide PKS in which the  
XX module 1 KS domain is inactivated by deletion or other mutation. In  
XX particular, the inactivation is mediated by a change in the KS domain  
XX that renders it incapable of binding substrate (the KS1-o mutation),  
XX oleandolide PKS is useful for synthesizing polyketides, which are useful  
XX as antibiotics and motifs. Heterologous expression of oleandolide PKS  
XX in host cells such as Streptomyces coelicolor and S. lividans is also  
XX made possible. Unmodified oleandolide compounds can be provided to  
XX cultures of Saccharopolyspora erythraea and converted to the  
XX corresponding derivatives of erythromycins A-D.  
XX Sequence 50937 BP; 6672 A; 16253 C; 19272 G; 8740 T; 0 other;

alignment\_scores:  
Quality: 472.50 Length: 423  
Ratio: 1.898 Gaps: 12  
Percent Similarity: 58.865 Percent Identity: 34.043  
alignment\_block:  
US-09-724-797-36 x AAA09469 ..  
Align seg 1/1 to: AAA09469 from: 1 to: 50937  
4 AspAlaValThrAlaPheAspProThrAspAlaAspValAlaArgAspPr 20  
|||||  
46524 GAGCGGTACCCGCTTAC...CCGTTGAGCTGCGGACGCGCTGAGACT 46570  
|||||  
20 OlyProSerThrHisTrpLeuArgHisAspProValHisArg... 35  
|||||  
46571 CGACCCGACACGCCGACACTCCGCCGACGACGACGCGCTCCAGGGTGC 46620  
|||||  
36 .....GlyAlaHisArgValTrpTyrValSerArgPheAlaAsp 48  
|||||  
46621 GCCTGACCTACGGCGAGGACGAGCTGCTGCTGCTGCTGCTGCTGCTG 46670  
|||||  
49 ValArgAlaValLeuGlyAspGluArgPheAlaArgThrGlyLeuArg 65  
|||||  
46671 GCCGCTATCGTTGCGGAGCTCCGCTTCAAGCAGCGGCGCGC... 46715  
|||||  
65 gPheTrpThrAspLeuValGlyProGlyLeu...AlaGluIleValG 81  
|||||  
46716 .....ACGATCCGCGCACGCCGCGGATGTTCCACCCGCGCGAGC 46758  
|||||  
81 LyAspIleIleLeuPheGlnAspGluProAspHisGlyArgLeuArgGly 97  
|||||  
46759 CGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 46808  
|||||  
98 ValValGlyProAlaPheSerProSerAlaLeuArgArgLeuProVa 114  
|||||  
46809 CTGCTGGGCAAGGCTTCAAGGACGCGCGGAGGAGATGCGGCGCG 46858  
|||||  
114 IleIleGlyThrValAspAspLeuArgProAlaLeuAlaArgGly 131  
|||||  
46859 TGTCGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 46908  
|||||  
131 Ia...MetAspValValAspGluLeuAlaArgProLeuAlaLeuAla 146  
|||||  
46909 CACCGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 46958  
|||||  
147 ValLeuGlyLeuLeuGlyLeuProAlaAlaAspTrpGlyValAlaGly 163  
|||||  
46959 ATCTGCGCAACTGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTG 46991  
|||||  
163 gTrpSerArgAspValGlyArgThrLeuAsp... 173  
|||||  
46992 .....CGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 47034  
|||||  
174 ..ArgGlyAlaSerAlaGluAspMetArgArgGlyHisAlaIleAla 189  
|||||  
47035 CCGGCGTACCGCGCGGAGATACAGGG...GTCAGCAG 47072  
|||||  
190 GluPheAlaAspTyrValGluArgAlaLeuAlaArgArgArgGluGlu 206  
|||||  
47073 GACTTATGTTATCATGACGCGCTGCTGCTGCTGCTGCTGCTGCTG 47122  
|||||  
206 yGlyGluAspLeu...AlaLeuMetLeuAspAlaHisAspArgGlyL 222  
|||||  
47123 CACCGAGGACTCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 47172  
|||||  
222 eumetSerArgAsnGluIleValSerThrValValThrHeilePheThr 238  
|||||  
47173 ACTGACCAAGGCGAGATCGTAACATGAGGGGTAGCCTGCTCATCGG 47222  
|||||  
239 GlyHisGluThrValAlaSerGlnValGlyAsnAlaValLeuSerLeu 255



```

174 rgglValaserAlaIgluspmetArgArgGlyHisAlaAlaIleAlaIglu 190
768 .....GTACGGCGGAGGAGATGCAG.....GAGCGCGCGAGCAGGCC 806
191 PheAlaAspTyrValIgluArgAlaLeuAlaArgArgArgGlyGly 207
807 TACGCC...TACATGGCGACCTCATGCAGCCGACGCTCGAGAGAACCCAC 853
207 ygluspleuLeuAlaLeuMetLeuAspAlaHisAsp...ArgGlyLeuM 223
854 GCACGATCTGTGAGCGCGCTGTCACAGCCCGTGCACACAGCAGACTGCT 903
223 eSerArgAsnGluIleValSerThrValAlaThrPheAlaPheThrGly 239
904 TGTCCGACAGCAACTGCTCGACTGGCCATTGCTGCTGCTGCTGCGGGA 953
240 HisGluThrValAlaSerGlnValIgluAsnAlaValLeuSerLeuAla 256
954 TACGAGACAGCAGCACCAGATCGCGACTTCGTGATCTCCTGATGAC 1003
256 aHisProAspGlnLeuAspLeuLeuArgArgArgProAspLeuAlaG 273
1004 CCGACCGGAGCTGCTCGACAGCTACTGATCGCGCGAGCTGATTCCT 1053
273 InAlaValIgluGlyLeuArgTyrAspProSerValIgluSerAsnThr 289
1054 CCGAGTCGAGAGAGCTGACGGGTGGCGCGCTGGGGTGGAGACGGCG 1103
290 .....ArgGlnLeuAspValAlaValIgluLeuArgGlyArgGlyLeu 304
1104 TTCGCCCTGACGCGGTGGAGATGTCGCTGCGGGGCGGACCATCCG 1153
304 gArgAspAspValValValAlaValLeuAlaGlyAlaAlaAsnArgAspPro 321
1154 CGCGGGCGAGCGCGGTGCTGCTGCGACCGGTGCCGACCGGAGCAGC 1203
321 rGatGlyTyrAspArgProAspAspPheAspIleGluArgAspProValPro 337
1204 CCCAGTTTCGCCGACGCGGATCGACGTCGACCGACCGACCCCAATCAG 1253
338 SerMetSerPheGlyAlaGlyMetArgTyrCysLeuGlySerTyrLeuAl 354
1254 CATCTCGGCTTCGACACGCGGTCCACCATGCTCGGCTCTCTCTGCGC 1303
354 aArgThrGlnLeuArgAlaAlaVal...AlaAlaLeuAlaArgLeuProG 370
1304 CCGGGTGGAACTCCAGGTGCAGTGCAGGTCTCTCCAGGACTGCCAG 1353
370 IyLeuAlaGlyLeuGlyCysAlaSerAspAlaLeuAlaTyrGlnProArgThr 386
1354 GGATTCGCGCTGGGATTCGCCGAGCAGCAGCTCGGTGAGCGAGAGGATG 1403
387 MetPheArgGlyLeuAlaSerLeuProIleAlaPhe 398
1404 TTGCTGCTGCGCCACATCGAGCTGCGGCTGCTGG 1439

seq_name: /sids1/gc9data/geneseq/geneseq-emb1/NA2001A.DAT.AAI71296
seq_documentation_block:
ID AAI71296 standard; DNA; 2795 BP.
XX
XX AAI71296;
XX
XX 24-JAN-2002 (first entry)
XX
XX Streptomyces griseus RppA protein encoding DNA SEQ ID NO.1.
XX
XX Streptomyces griseus; RppA; pyran-2-one; drug; agricultural; ds.
XX
XX Streptomyces griseus.
XX
XX
XX Key Location/Qualifiers

```

```

FT CDS 1596..2714
FT /tag=a
FT /product="RppA protein"
XX JP2001238677-A.
XX
XX 04-SEP-2001.
XX
XX 01-MAR-2000; 2000JP-0055782.
XX
XX 01-MAR-2000; 2000JP-0055782.
XX
XX (MITR ) MITSUBISHI RAYON CO LTD.
XX
XX WPI: 2001-613891/71.
XX P-PSDB: AAG68148.
XX
XX Preparation of pyran-2-one derivative, used as intermediate and as
XX additive for thermosensible transfer paper, comprises contacting
XX carbonyl compound with microbe containing polyketide synthase gene -
XX Claim 6; Page 7-8; 11pp; Japanese.
XX
XX The present invention describes the preparation of pyran-2-one
XX derivatives. The preparation method comprises contacting carbonyl
XX compounds with microbes containing a polyketide synthase gene.
XX Pyran-2-one derivatives can be used as synthetic intermediates for
XX drugs and agricultural chemicals. They can also be used as additives
XX for thermosensible transfer paper. The present sequence encodes the
XX specifically claimed Streptomyces griseus RppA protein, which is used
XX in the exemplification of the present invention.
XX
XX Sequence 2795 BP; 435 A; 1051 C; 908 G; 401 T; 0 other;

alignment_scores:
Quality: 461.00 Length: 412
Ratio: 1.837 Gaps: 15
Percent Similarity: 60.922 Percent Identity: 33.495

alignment_block:
US-09-724-797-36 x AAI71296 ..
Align seg 1/1 to: AAI71296 from: 1 to: 2795

16 ValArgArgAspProTyr.....Prose 23
374 GTCCGGAACGTGCTTCGACTACGCGCAGCAGCTGAGTGCACCCCA 423
23 rTyrHisThrPleuLeuArgHisAspProValHisArg..... 35
424 GCTCAGGCAATGCTCACCAGGAGCGGCTGTCGCCATCCGTATGGCGT 473
36 ..GlyAlaHisArgValTyrTyrValSerArgPheAlaAspValAla 51
474 ACGAGAGAGGCGAGCGCTGCTGTCACCCGCTACGAGAGAGCTCGGAGC 523
52 ValLeuGlyAspGluArgPheAlaArgThrGlyIle.....ArgArgP 66
524 GTCCACCACCGACCGCGGCTTCAGCCGACGCGCTGCTCGCGGAGACTT 573
66 eTrpThrAspLeuValGlyProGlyLeuLeuAlaGlu.....IleValG 81
574 C.....CCCCGATGACCCCGAGCCGATGCGCAG 605
81 IyAspIleIleLeuPheGlnAspGluProAspHisGlyArgLeuArgGly 97
606 CGGAGTCATCAACCTCATGAGCAGCCGCGCAGACCGGCTGCGGAGC 655
98 ValValGlyProAlaPheSerProSerAlaLeuAlaGlyLeuGluProVa 114
656 CTGGTGGCCAGAGCTTCACCCG.....CGTGGCTGAGACAGAT 696

```





Ratio: 1.814 Gaps: 13  
Percent Similarity: 59.390 Percent Identity: 30.282

alignment\_block:

US-09-724-797-36 x AA014548 ..

Align seg 1/1 to: AA014548 from: 1 to: 1400

```

11 ProthAspAlaSpValArgArgProTyrPro..... 22
137 CTTACCCGAGTGAAGTGTATGACGACCGTTCGGATCTCGAAAGCGA 186
23 .SeryrHis.....Trrleu.....LeuArgHis.... 30
187 CTCCTTCACGTCGACTGTACCGGACCTACCCGACGCTGCCGAGACCG 236
31 .....AspProValHisArgGlyAlaHisArgValTrrTyrValSer 44
237 CCGCGGTGACCGCGGTGCTTCGCGGACGAGGACGCTGGCTGTACAC 286
45 ArgPheAlaSpValArgAlaValLeuGlyAspGluArgPheAlaArgTh 61
287 GGCTACGACGAGCGAAGCGCGCTGAGCGACTGCGCTGACGACGGA 336
61 rGlyIleArgArgPheTrrPThrAspLeuValGlyProGlyLeuAlaG 78
337 CCGGAAGAAGATGAC.....CCGGCGCTGGAGGTGCG 368
78 IuIleValGly.....AspIleIleLeuPhe..... 86
369 AGTTCGCGCATACCTCGGTTTCCCGGAGGACGTGCGAACTACTTCC 418
87 .....GluAspGluProAspHisGlyArgLeuArgGlyVal 98
419 ACCAACATGGGACACGAGGACCGCGACCCACACCGCGCTGCGCACT 468
98 IValGlyProAlaPheSerProSerAlaLeuArgArgLeuProValI 115
469 GGTGCGGAGAGTTCACCGTCCGCGGTGAGGCGATGGCGGCCCGCG 518
115 IeAlaGlyThrValAspAspLeuLeuArgProAlaLeuAlaArgGly 131
519 TCGACACAGATCCCGGAGCTGTCGACGAGTGGGCGACTCGGCGCTG 568
132 MetAspValAlaSpGlyLeuAlaAlaAspTrrPglyAlaValGlyArg 165
569 GTGACATCGCTACCGCTCCGACCCGCTGCCCATCAAGTCATCTG 618
148 uGlyLeuLeuGlyLeuProAlaAlaAspTrrPglyAlaValGlyArg 165
619 CGAGCTGCTCGGCTGCGACGAGAAATACCGGCGGAGTTCGGCGGTGA 668
165 eArgAspValGlyArgThrIleuAspArgGlyAlaSerAlaGluAspMet 181
669 GCTCGGAGATC...CTGTCATGACCCGCGGAGCGGCGGAACAG.... 709
182 ArgArgGlyHisAlaAlaIleAlaGlyPheAlaAspTrrValGlyArg 198
710 ...CGCGGCGAGGCGGCC...AGGAGGTGCTGCAATTCATCTCGACT 753
198 AleuAlaArgArgArgGlyGlyGlyAlaAspLeuAlaLeuMetL 215
754 GGTGAGGCGCGCGCACCGACCGCGGAGACCTGTGTCGCGGCTTA 803
215 euAspAlaHis.....AspArgGlyLeuMetSerArgAspGluIleVal 229
804 TCAGGGTCCAGGACGACGATGACGTGCTGCTACGCGCGGAGAGTGTCC 853
230 SerTrnValValThrPheIlePheThrGlyHisGluThrValAlaSerG 246
854 TCATCGCGCGCTGTGCTGTGCTGCGCGGTTCGAGGCGGTGAGGCT 903
246 nValGlyAlaAlaValLeuSerLeuAlaHisProAspGlnLeuAspRL 263

```

```

904 CATGGGATCGGCGACCTACTGCTGCTACCCACCGCGGACCTGCGCGC 953
263 euleuArgArgArgProAspLeuAlaGlnAlaValGluGlyLeu 279
954 TGTGCGGCGGAGACCGTCCGCTGCGCAACCGCGCTGAGAGATCTG 1003
280 ArgTyrAspProSerValGlnSerAsnThrArgGlnLeuAspValAsp 296
1004 CGCTACATCGCTGACGAGGACACCGCGCTGCGCGGAGAGACCT 1053
296 IGluleuArgGlyArgArgLeuArgArgAspValValValLeu 313
1054 GGNATGCGGCGTCTCCGATCCCGCATGACACGCGTCTGTGCGCA 1103
313 IagIyAlaAlaAsnArgAspProArgArgTyrArgAspArgProAspPhe 329
1104 ACGCGCGGCGCAACCGGACCGAAGCATGTTCCGCGGACCCGCTCC 1153
330 AspIleGluArgAspProValProSerMetSerPheGlyAlaGlyMetAr 346
1154 GACGTACCGCGGACACCGCGGCCACCTGTGCTGCGGACGAGCATCA 1203
346 gTyrCysLeuGlySerTyrLeuAlaArgThrGlnLeuArgAlaValA 363
1204 CTTCGTGATGGCGCGCGCTGGCGCAAGCTGAGGCGAGAGTGGCTGC 1253
363 IaAlaLeu...AlaArgLeuProGlyLeuArgLeuGlyCysAlaSerAsp 378
1254 GGGCGCTGTCGCGCGCTGCCGCTGTGCTGCGGAAATCGACGCGGAC 1303
379 AlaleuAlaTrgInProArgTrnMetPheArgGlyLeuAlaSerLeuP 395
1304 GACGTGCTGTGGCGGCTTGGGTGCTGTGCGGGGATGACCACTTACC 1353
395 oIleAlaPhe.....ThrProGly 401
1354 GGTGCGGCTGACGAGATGACACCTGGC 1381

seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AAA95667
seq_documentation_block:
ID AAA95667 standard; DNA; 1204 BP.
XX
AC AAA95667;
XX
DT 14-FEB-2001 (first entry)
XX
DE Bacillus subtilis hydroxylating enzyme gene #2.
XX
KW Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor; ds;
KW hydroxymethylglutaryl co-enzyme A reductase; hypocholesterolaemia.
XX
OS Bacillus subtilis.
XX
PN WO20044886-A1.
XX
PD 03-AUG-2000.
XX
PF 28-JAN-2000; 2000WO-JP00472.
XX
PR 29-JAN-1999; 99JP-0021707.
XX
PA (KYO) KYOMA HAKKO KOGYO KK.
XX
PI Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;
XX
DR MPI: 2000-548827/50.
XX
DR P-PSDB; AAB15502.
XX
PT New protein derived from Bacillus genus microorganism useful for
PT producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase
PT inhibitors

```



...





```

5543 CGACACATTCGGAGCTGTTCACTCATCTATATAAGGAGTTCTTAAAT 5592
38 1sargvalttrptryalserarphhealaaspyalaargalavaluengly 54
5593 ACCCGGCTGGATGTACAGGATATGAGAAAGCGCTGCTATTTTGA 5642
55 aspgluatgphe...Alaargthrglyleargarphetrphtasphe 70
5643 GATCGGAGATTCAAAAGTCCGACCCCGCTGCTGAGACCTCAACCAATA 5692
70 uValoJyproglyleuLeuAlaGuIleValGlyAspIlelleueupheg 87
5693 TCAGGAC.....CTTTCACATGTCAAAATCAATGATGCTGTTTC 5733
87 1nAspGluProAspHisglYarGleuAargLyValaIGlyProAlaphe 103
5734 AGAACACACCTGATCATAGACGATTGGGACGCTTGCCAGCGGACGTTT 5783
104 SerProSerAlaLeuAargArgLeuGluProValIlealaglythValas 120
5784 ACGCCGAGAACGACGAGAGATTATCAGCCGTATATCATTTGAACGTCCA 5833
120 paspleuAargrproAlaLeuAlaAargGlyAlaMetaspValaAspG 137
5834 TCATTGTCTTATCAAGTCAAGTAAAGTAAAGATGAGGTCAATTTCGG 5883
137 1uLeuAlaTyPProleuAlaLeuAargAlaValleuGlyLeuenglyLeu 153
5884 ACTTGCTTTTCTTACCAAGTTTGTTCATAGCTAACATATATAGTCTA 5933
154 ProAlaAlaAspTrpGlyAlaValaGlyArGTTrpSerArGAspValGlyAr 170
5934 CCGGAGGAAGATAGGAGCAATTAAGAGATGGCGCTCGCATCTCATCA 5983
170 gPhrleuAspArgLyAlaSerAlaGluAspMetArGArgGlyHisAla 187
5984 AACGATGATTTTACCCTCTCAAGAAAGCATTAACAGAGGCAATATTA 6033
187 1aIleAlaGluPheAlaAspTyValaIGuarAlaLeuAlaAargArg 203
6034 TGGCGTGCAGGCTATGGCATATTTCAAAAGCGATTCAAAGGAAAA 6083
204 ArgGluGlyGlyAspLeuAlaLeuMetLeuAspAlaHisAspAr 220
6084 CGCCACCTCAACAGATATGATCAGATCCTCTTGAAGGAGAGAAAA 6133
220 gGlyLeuMetSerArGAsnGluIleValSerThValaValhrheilep 237
6134 GGATTAAGCTGAGGAAGAGAGCGCGCATCTACGCTATTTGCTGGCGA 6183
237 heThrGlyHisGluThrValaAlaSerGlnValaGlyAsnAlaValleuSer 253
6184 TCGCCGAGACATGAGCAACGGCATCTCATCAGCAATTCAGTCTTTGT 6233
254 LeuLeuAlaHisProAspGlnLeuAspLeuAargArGArgProAspLe 270
6234 CTGCTGCAGCATCCAGAACAGCTTTTGAACGTAGAGAAATCCAGATCT 6283
270 uLeuAlaGlnAlaValaIGluGlyLeuAargTyArGAspProSerValGlns 287
6284 TATTTGGTACCGAGTCCGAGAAATGTTTACGCTATGAAAGCCCGCAGAAA 6333
287 eArAsnThrArgGlnLeuAspValaIGluLeuAargLyArGArgLeu 303
6334 TGACAGCCAGAGTTGCTCAGAGATATTGACATCTGCGGGGTGACGATC 6383
304 ArgArGAspAspValaValaValleuAlaGlyAlaAlaAsnArGAspPr 320
6384 CGTCAAGAGAAAGAGCTATCTTTTGTAGAGGCGCTATTCGAGACCC 6433
320 CAArgArGtyrAspArgProAspAspPheAspIleGluArGAspProValP 337
6434 TAGCATATTTCAGAACCCCGATGCTCTGATATTACGAGAAAGTCTTAATC 6483

```

```

337 roserMetSerPheGlyAlaGlyMetArGtyrCysLeuGlySerTyLeu 353
6484 CGCATCTTTTATTCGGGATGGCATCATGTTTGTCTAGGTCCTCGCG 6533
354 AlaArgThrglnLeuAargAlaVala...AlaAlaLeuAlaAargLeuDr 369
6534 GCACGATTAGAGCGCAAAATTCGATTAAACACTCTTCTCAGCGAATGCC 6583
369 oGlyLeuAargLeuGlyCysAlaSerAspAlaLeuAlaTyrglnProArgT 386
6584 CAGCCTTAATCTT.....GCGGATTTTGATGCGCGGTATCGCGCTTT 6627
386 hMetPheArgGlyLeuAlaSerLeuProIleAlaPhe 398
6628 TTGGATTTGCGGCGCTTGAGAGACTGCCGCTGACTTTT 6665

```

```
seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-407-549-1
```

```

seq_documentation_block:
; Sequence 1, Application US/09407549
; Patent No. 6303377
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley Grant
; APPLICANT: Perkins, John B.
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Petro, Janice G.
; TITLE OF INVENTION: BIOTIN BIOSYNTHESIS IN BACILLUS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/407, 549
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/239, 430
; FILING DATE: May 6, 1994
; APPLICATION NUMBER: 08/084, 709
; FILING DATE: June 25, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29, 066
; REFERENCE/DOCKET NUMBER: 04599/004001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8478
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-09-407-549-1

```

```

alignment_scores:
Quality: 609.00 Length: 396
Ratio: 2.316 Gaps: 5
Percent Similarity: 66.414 Percent Identity: 33.586

```

alignment\_block:  
US-09-724-797-36 x US-09-407-549-1 ..

Align seg 1/1 to: US-09-407-549-1 from: 1 to: 8478

```

8 AlaphespprothrAspAlaaspValArgarGspProtyrProserTy 24
  ||| : : : : : : : : : : : : : : : : : : : : : : : :
5493 GCATGCTCAACTGCATCTTCTGAGTTTGGAAACCATTATCTTTT 5542
24 HistripleuLeuArgHisAspProvalHisArgGlyAla.....H 38
  ||| : : : : : : : : : : : : : : : : : : : : : : : :
5543 GCACACATTGCGAGCTGTTTCATCTCATTAAGAGGAGTTTCTTAAT 5592
38 iArgValIrrptIvalSerArgPheAlaaspValArgAlaValLeuLy 54
  : : : : : : : : : : : : : : : : : : : : : : : : : :
5593 ACCCGGCTGATATGTCACAGGATATGAGAAACGGCTGTATTTTAAA 5642
55 AspGluArgPhe...AlaArgThrGlyIleArgArgPheTrpThrAsp 70
  ||| ||||| : : : ||||| : : : : : : : : : : : :
5643 GATGGAGATTCAAGTCCGACCCCGCTGAGAGCTCACCAATA 5692
70 uValIlyProGlyLeuLeuAlaGluIleValGlyAspIleIleuPheG 87
  ||| : : : : : : : : : : : : : : : : : : : : : : : :
5693 TCAGGAC.....CTTCACATGTCAAAATCAAAATGATGCTGTTTC 5733
87 InAspGluProAspHisGlyArgLeuArgGlyValIleGlyProAlaPhe 103
  ||| : : : ||||| ||||| ||||| : : : : : : : : : :
5734 AGAACACCGCTGATCATAGACGATTGCGAGCGTTGCCAGCGGAGCTT 5783
104 SerProSerAlaLeuArgArgLeuGluProValIleAlaGlyThrValas 120
  : : : : : : : : : : : : : : : : : : : : : : : :
5784 ACGCCGAGACGACAGAGATTATCAGCCGTATATCATTTGAAGACTGCCA 5833
120 PaspleuLeuArgProAlaLeuAlaArgGlyAlaMetAspValAspG 137
  ||||| : : : : : : : : : : : : : : : : : : : : : :
5834 TCATTGCTTATCATCAGTCAGCAAGGAAAAAAAAGATGGAGCTCATTTCCG 5883
137 IuLeuAlaTyrrProLeuAlaLeuArgAlaValLeuGlyLeuGlyLeu 153
  : : : : : : : : : : : : : : : : : : : : : : : :
5884 ACTTGCTTTTCCTTTAGCAAGTTTTCATACGTAACATATATAGTGTA 5933
154 ProAlaAlaAspTrpGlyAlaValGlyArgTrpSerArgAspValGlyArg 170
  ||| : : : : : : : : : : : : : : : : : : : : : :
5934 CCGGAGGAGATAGGAGCAATTAAGAGAGTGGCTGCGATCTCATTTCA 5983
170 gThrLeuAspArgGlyAlaSerAlaGluAspMetArgArgGlyHisAla 187
  : : : : : : : : : : : : : : : : : : : : : : : :
5984 AACGATTGATTTTACCGCTCAAGAAAGGCTTTAACAGAGGCAATATTA 6033
187 IalIleAlaGluPheAlaAspTyrValGluArgAlaLeuAlaArgArg 203
  : : : : : : : : : : : : : : : : : : : : : : : :
6034 TGGCTGTGACAGCTATGCGCATATTTCAAGAGCTGATTCAAAGAGAAA 6083
204 ArgGluGlyGlyIuAspLeuLeuAlaLeuMetLeuAspAlaHisAspArg 220
  ||||| : : : : : : : : : : : : : : : : : : : : : :
6084 CGCCACCTCAACAGAGATATGATCAGCATGCTTTGAAGGGGAGAGAAA 6133
220 gGlyLeuMetSerArgAsnGluIleValSerThrValValIrrPheIleP 237
  : : : : : : : : : : : : : : : : : : : : : : : :
6134 GGATTAAGTGCAGAGAGAGGAGGCGCATCTACGTCATATGCTGGCGA 6183
237 heTrpGlyHisGlyThrValAlaSerGluValGlyAsnAlaValLeuSer 253
  : : : : : : : : : : : : : : : : : : : : : : : :
6184 TCGCGCGCATAGAGCAACGGTCAATCTCATACCAATTCAGTCTTTGT 6233
254 LeuLeuAlaHisProAspGluLeuAspLeuLeuArgArgArgProAsp 270
  ||||| ||||| : : : ||||| : : : : : : : : : :
6234 CTGCTGCAGCATCCAGAACGCTTTTAAATCGAGAGAAATCCAGATCT 6283
270 uLeuAlaGluAlaValGluGlyCysLeuArgTyrAspProSerValGlns 287
  ||| : : : ||||| ||||| ||||| : : : : : : : : : :
6284 TATTGGTACCGCAGTCCAGAGATTTTACGCTATGAAGCCCAAGCAAA 6333

```

```

287 eRAsnThrArgGluLeuAspValAspValIuLeuArgGlyArgArgLeu 303
  : : : : : : : : : : : : : : : : : : : : : : : :
6334 TGACAGCCAGAGTTGCGTCAGAGGATATGACATCTGCGGGGAGCATC 6383
304 ArgArgAspAspValValValLeuAlaGlyAlaAlaHisAsnArgAspTr 320
  ||||| : : : ||| : : : ||||| ||||| ||||| |||||
6384 CGTAAAGAGAGAACAGTCTATCTTTGTGTAGAGAGCGGTATATGAGACC 6433
320 oArgArgTyrAspArgProAspAspPheAspIleGluArgAspProValP 337
  : : : : : : : : : : : : : : : : : : : : : : : :
6434 TAGCATATTCACAGAACCCGATGCTTCGATATTTACAGAAAGTCATATC 6483
337 roSerMetSerPheGlyAlaGlyMetArgTyrCysLeuGlySerTyrLeu 353
  || : : : ||||| ||| : : : ||||| ||||| |||
6484 CGCATCTTTCATTCGGGCATGGCCATCATGTTTGGCTTAGGGTCTCGCTG 6533
354 AlaArgThrGluLeuArgAlaAlaVal...AlaAlaLeuAlaArgLeuP 369
  ||||| : : : : : : : : : : : : : : : : : :
6534 GCACGATTAAAGCGCAAAATTCGATTAACACTTCTCGAGCGAATGCC 6583
369 oGlyLeuArgLeuGlyCysAlaSerAspAlaLeuAlaTyrGluProArgT 386
  : : : : : : : : : : : : : : : : : : : : : : : :
6584 CAGCCTTAATCTT.....GCGGATTTTGAATGCGCGTATCGCGCTTT 6627
386 hMetPheArgGlyLeuAlaSerLeuProIleAlaPhe 398
  ||||| : : : : : : : : : : : : : : : : : :
6628 TTGGATTTCGGGCGCTTGAGAGAGCTGCCGCTGACTTTT 6665
seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:US-09-105-537-38
seq_documentation_block:
; Sequence 38, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.4380S1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 38
; LENGTH: 1251
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; US-09-105-537-38
alignment_scores:
Quality: 538.50 Length: 398
Ratio: 2.216 Gaps: 8
Percent Similarity: 61.055 Percent Identity: 34.673
alignment_block:
US-09-724-797-36 x US-09-105-537-38 ..
Align seg 1/1 to: US-09-105-537-38 from: 1 to: 1251
15 AspValArgArgAspProtyrProserTyHisTrpLeuAlaHisAs 31
  ||| : : : : : : : : : : : : : : : : : : : : : :
67 GATTTGGCGCGCGATCCGATATCGAGCTACGAGACTGCGTGGCGAGGG 116
31 pProValHisArg.....GlyAlaHisArgValIrrTyrValas 44
  ||||| ||||| : : : ||||| : : : : : : : : : :
117 TCGGCGCCACCGGATCGCAGCCCGGAGGAGAGAGTGTGGCGTGTG 166
44 eArgPheAlaAspValAlaValLeuGlyAlaAspIuArgPheAlaArg 60
  : : : : : : : : : : : : : : : : : : : : : : : :
167 TCGGCTACGACCGGCGCGGCTTCCTCGCGATCCCGGTTTCAGCAAG 216

```

```

61 ThrGlylleArgArgPheTrpThrAspLeuValGlyProGlyLeuLeuAl 77
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
217 GAC.....TGGCGCAACTCCACGACACTCCCTGACGAGGC 251
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
77 aglu...lIeValGlyAspIleIleLeuPheGlnAspGluProAspHisG 93
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
252 CGAGGCGCGCTCAACCAACATGCTGTGAGTCCGACCCCGCGGCACA 301
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
93 lyaArgLeuArgGlyValValGlyProAlaPheSerProSerAlaLeuArg 109
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
302 CCGCGCTGCGCAAGCTGTGTGGCCGTGATCCATCGCCCGGTTGAG 351
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
110 ArgLeuGluProValIleAlaGlyThrValAspAspLeuArgProAl 126
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
352 TTGCTGGGCGCCCGGCTCCAGAGATGCTGACGGGCTCGTGAGCGCAT 401
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
126 aleuAlaArg.....GlyAlaMetAspValValAspGluLeuAlaTrp 141
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
402 GCTGGCGGCGCCGCGCGCGCGCATGTGATGAGTCCCTGGCGTGGC 451
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
141 roLeuAlaLeuArgAlaValLeuGlyLeuGlyLeuProAlaAlaAsp 157
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
452 CGGTGCGGATCACCGTATCTCCGAACTCTCGCGGTGCCGAGCCGAG 501
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
158 TrpGlyAlaValGlyArgTrpSerArgAspValGlyArgThrLeuAsp 174
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
502 CGCGCGCGCTCCCGCTGTGACC.....GACGC 530
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
174 gGlyAlaSerAlaGluAspMetArgArgGlyHisAlaAlaIleAlaGlu 191
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
531 CTGTGTTCCCGGACGATCCCGCCGAGGCCAGACCGCATGGCCGAGA 580
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
191 heAlaAspTrpValGluArgAlaLeuAlaArgArgArgGluGly 207
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
581 TGAGCGCTATCTCTCCCGCTCATGCACTCCAAAGCGGGGAGAGAGGC 630
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
208 GluAspLeuLeuAlaLeuMetLeuAspAlaHisAspArg.....GlyLe 222
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
631 GAGGACCTGCTCAGCGCGCTGTGCGGACGACGAGGAGCGGCTCCG 680
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
222 uMetSerArgAsnGluIleValSerThrValValThrPheIlePheThrG 239
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
681 GCTGACCTCCGAGGAGCTGCTGATGCGCACATCCTGCTCTCCGCG 730
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
239 lYHisGluThrValAlaSerGlnValGlyAlaValLeuSerLeuLeu 255
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
731 GGCAGAGACCAAGCTCATCTGATCGCAACGGCATGTACGCGCTGCTC 780
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
256 AlaHisProAspGlnLeuAspLeuLeuArgArgArgProAspLeuAl 272
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
781 TCGACGCCGACAGCTGCGCGCTGGGCGCACATGACGCTCTTGA 830
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
272 aglAlaValalGluGlyCysLeuArgTrpAspProSerValGlnSerSnt 289
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
831 CGGCGCGGTGAGGAGATGTTGCGCTACGAGGCCCGGTGAATCCGGA 880
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
289 hrArgGlnLeuAspValAsp...ValGluLeuArgGlyArgArgLeuArg 304
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
881 CCTACCGCTTCCCGCTGACGCCGCTGACACCTGACGAGGAGCGTATCCG 930
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
305 ArgAspAspValValValLeuAlaGlyAlaAlaAsnArgAspProAr 321
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
931 GCGGCTGACAGGTCCTGCTGCTGCGACGCCACCGCACCCGCA 980
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
321 gArgTrpArgArgProAspAspPheAspIleGluArgArgProValPro 338
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
981 GCGCTTCCGAGACCGCACCGCTTGCATCCGCGGAGACACCGCGGCC 1030
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
338 erMetSerPheGlyAlaGlyMetArgTrpCysLeuGlySerTrpLeuAl 354
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1031 ATCTGCGCTTGGCCAGCGCATCCACTTCTGCAATCGCGCCCTTGCC 1080
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
355 ArgThrGlnLeuArgAlaAlaVal...AlaAlaLeuAlaArgLeuProG 370

```

```

1081 CGGTGGAGGCCGGGATGCGCTCCGCGCCCTTCTCGAACCGTGCAGGA 1130
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
370 yLeuArgLeuGlyCysAlaSerAspAlaLeuAlaTrpGlnProArgThrM 387
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1131 CCTCGCGCTGACAGCTCTCCCGCGCAACTCTGTGTGATCCGAACCCGA 1180
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
387 etPheArgGlyLeuAlaSerLeuProIleAlaPheThrProGly 401
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1181 TGATTCGCGGCTCAAGGCCCTGCGGATCCGCTGGCGGAGAGA 1224
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

seq_name: /cgn2_6/prodata/1/ina/6A.COMB.seq:US-09-320-878-21

seq_documentation_block:
; Sequence 21, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT MARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; EARLIER FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 5970
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-21

alignment_scores:
      Quality: 538.50      Length: 398
      Ratio: 2.216      Gaps: 8
      Percent Similarity: 61.055      Percent Identity: 34.673

alignment_block:
US-09-724-797-36 x US-09-320-878-21 ..
Align seg 1/1 to: US-09-320-878-21 from: 1 to: 5970

15 AspValaArgArgAspProGlyTrpProSerTrpHisTrpLeuLeuArgHisAs 31
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1422 GATTTCGCGGCGCATTCGTATCCGAGTACGCGACAGCTCGTGCAGAGG 1471
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
31 pProValHisArg.....GlyAlaHisArgValITrpyValS 44
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1472 TCCGCGCCACACGGGTGCGCACCCCGAGGGGAGCGAGGTGTGCTGTG 1521
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
44 erArgPheAlaAspValArgAlaValLeuGlyAspGluArgPheAlaArg 60
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1522 TCGGCTACGACCGGCGCGCGGTCTGCGCATCCCGGTTCAGACAG 1571
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
61 ThrGlylleArgArgPheTrpThrAspLeuValGlyProGlyLeuLeuAl 77
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1572 GAC.....TGGCGCAACTCCACGACACTCCCTGACGAGGC 1606

```

```
77 aglu...ilevalGlyAspIleIleuPhGlnAspGluProAspHisG 93
||||| :||| :||| :||| :||| :|||
1607 CGAGGGCCGGCTACACACATGCTGATGCCAGCCGCCGCCACA 1656
93 LysArgLeuArgGlyValValGlyProAlaPheSerProSerAlaLeuArg 109
||||| :||| :||| :||| :||| :|||
1657 CCGCGCTGGCGCAAGCTGTGGCCCGTAGTGCATGCATGGCCGGGTGAG 1706
110 ArgLeuGluProValIleAlaGlyThrValAspAspLeuArgProAl 126
||||| :||| :||| :||| :||| :|||
1707 TTGCTGGGGCCCGGGTCCAGAGATCGTCGAGGGCTCTGTGAGCCAT 1756
126 AluAlaIAspG...GlyAlaMetAspValValAspGluLeuAlaTyrP 141
||||| :||| :||| :||| :||| :|||
1757 GCTGGCGGCGCCCGAGCGCCGCGCTGATGATGATCCCTGGCTGGC 1806
141 roleuAlaLeuArgAlaValIleuGlyLeuLeuGlyLeuProAlaIAsp 157
||||| :||| :||| :||| :||| :|||
1807 CGCTGGCGGATCACCCTGATCTCCGAATCTCTGGCGTGGCCGACGCCGAG 1856
158 TrpGlyAlaValIleArgTyrPserArgAspValGlyArgThrLeuAspAr 174
||||| :||| :||| :||| :||| :|||
1857 CGGCGCGCTTCCTCCGCTGAGCC...GAGCG 1885
174 gGlyAlaSerAlaGluAspMetArgArgGlyHisAlaIleAlaGluP 191
||||| :||| :||| :||| :||| :|||
1886 CTTCGCTTCCCGGAGATCCGCCGAGGCCGACGCGCATGGCCGAGA 1935
191 heAlaAspTyrValGluArgAlaLeuAlaArgArgArgGlyGly 207
||||| :||| :||| :||| :||| :|||
1936 TGAGCGGCTTCCTCCGCTCATGACTCCAGCGCGGAGGAGCAGCG 1985
208 GlnAspLeuLeuAlaLeuMetLeuAspAlaHisAspArg...GlyLe 222
||||| :||| :||| :||| :||| :|||
1986 GAGGACTGCTGACCGCGCTCGCGGAGCAGCAGCAGAGAGGCGTCCG 2035
222 uMetSerArgAsnGluIleValIleSerThrValValThrPheIlePheThrG 239
||||| :||| :||| :||| :||| :|||
2036 GCTGACCTCCGAGGACTGCTCGTGTGGCCGACATCTCTGCTGGCG 2085
239 LysHisGluThrValAlaSerGlnValGlyAsnAlaValLeuSerLeu 255
||||| :||| :||| :||| :||| :|||
2086 GGCAGGAGACACGATCATCTGATGCCAGCGCATGACGCTGCTC 2135
256 AlaHisProAspGlnLeuAspLeuLeuArgArgArgProAspLeuAla 272
||||| :||| :||| :||| :||| :|||
2136 TCGACCCCGACAGCTGGCGCGCGCGCGCAGCATGACGCTCTTGA 2185
272 aglnAlaValIleGluIleuSerLeuArgTyrAspProSerValGlnSerAsnT 289
||||| :||| :||| :||| :||| :|||
2186 CGGCGCGCTGAGAGATGTGGCTACGAGGCGCGCGGATCCGCA 2235
289 hrArgGlnLeuAspValAsp...ValGluLeuArgGlyValArgLeuArg 304
||||| :||| :||| :||| :||| :|||
2236 CTTACCCCTTCCTCCGCTGAGCCCGCTGACGCGCATGCCG 2285
305 ArgAspAspValValValValIleuAlaGlyAlaAlaAsnArgAspProAr 321
||||| :||| :||| :||| :||| :|||
2286 GCGGCTGACAGGCTCTCTGCTGCTGGCGGAGCGCCAGCAGCCCGCA 2335
321 gArgTyrAspArgProAspAspPheAspIleGluArgAspProValProS 338
||||| :||| :||| :||| :||| :|||
2336 GCGCTTCCCGAGCCCGACCGCTTCGACATCCGCGGAGACCGCGCGC 2385
338 erMetSerPheGlyAlaGlyMetArgTyrCysLeuGlySerTyrLeuAla 354
||||| :||| :||| :||| :||| :|||
2386 ATCTCGGCTTCGGCGACGCGATCTCTGATCGCGCGCCCTTGGCC 2435
355 ArgThrGlnLeuArgAlaAlaVal...AlaAlaLeuAlaArgLeuProS 370
||||| :||| :||| :||| :||| :|||
2436 CGGTTGGAGGCCGAGATCGCGCTTCGCGCTTCGAGACGCTCGCGGA 2485
370 yLeuArgLeuGlyCysAlaSerAspAlaLeuAlaTyrGlnProAlaGlyThm 387
```

```
2486 CCTCGCCCTGAGCTCTCCCGCGGCAACTGCTGTGATCCGACCGA 2535
387 etPheArgGlyLeuAlaSerLeuProIleAlaPheThrProGly 401
||||| :||| :||| :||| :||| :|||
2536 TGATCCGCGGCTCAAGGCGCTCGCATCCGCTGGCGCGAGGA 2579
seq_name: /cgn2_6/ptodata/1/fna/6B_COMB.seq:us-09-029-603-4
seq_documentation_block:
; Sequence 4, Application US/09029603
; Patent No. 6210935
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Engel, Natalie
; APPLICANT: Bietenader, Jurg
; APPLICANT: Toupet, Christine
; APPLICANT: Pospiech, Andreas
; TITLE OF INVENTION: Staurosporin Biosynthesis Gene Clusters
; FILE REFERENCE: 4-2055/A/PCT
; CURRENT APPLICATION NUMBER: US/09/029,603
; EARLIER FILING DATE: 1998-03-20
; EARLIER APPLICATION NUMBER: PCT/EP96/03643
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 6085
; TYPE: DNA
; ORGANISM: Streptomyces longisporoflavus
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: (378)..(1665)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: (1747)..(2553)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: (2593)..(4011)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: (4013)..(4999)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc_RNA
; LOCATION: (5071)..(6085)
; OTHER INFORMATION: ORF
US-09-029-603-4
alignment_scores:
Quality: 508.50 Length: 407
Ratio: 2.101 Gaps: 11
Percent Similarity: 59.459 Percent Identity: 36.364
alignment_block:
US-09-724-797-36 x US-09-029-603-4 ..
Align seg 1/1 to: US-09-029-603-4 from: 1 to: 6085
20 ProTyrProSerTyrHisThrLeuAlaArgHisAspProValHisArg.. 35
||||| :||| :||| :||| :||| :|||
441 CCTTACCCGCTTACCGGCGGCTACCGGAGCGCGCCCGCTCATCGAC 490
36 .....GlyAlaHisArgValTyrTyrValSerArgPheAla 48
||||| :||| :||| :||| :||| :|||
491 GCGCGCGGCGCCGGAACCTGACACCTACTACTGCTTCACTACGACG 540
48 sPValArgAlaValIleuGlyAspGluArgPheAlaArgThrGlyLeuArg 64
||||| :||| :||| :||| :||| :|||
```

```

541 ACGTGGTCCGGCTCTGTCCACCGCGCGTTCGCGCCGACACGCC..... 584
65 ArgPheThrPThrAspLeuValGlyProGly..... 74
585 CGGTGGCTCTCCGGGACACCGCGCGCCGACACCGCGCGCTCCGATCCC 634
75 .....LeuLeuAlaGluIleValGlyAspIleLeuPheG 87
635 GCGCGAGACCGCGCGCTGGGACCGCTGTCGAGAACGTGGCTGTCTTCC 684
87 LAspGluProAspHisGlyArgLeuArgGlyValGlyProAlaPhe 103
685 TCGACCCCGCGCGCCACACCACTGCGCTCCCTCCACCGCGCGAGTTC 734
104 SerProSerAlaLeuArgArgLeuGluProValIleAlaGlyThrValAs 120
735 TCACCCCTGATGTCACCGCGCTGGCGCCCGCATCGCGAACTCGCGAG 784
120 PAspLeuLeuArgProAlaLeuAlaArgGlyAlaMetAspValAlaSpG 137
785 CGAACTCTTGACCGCGCTCGAGACACCGCGCGCGCATCTCTGTCGAGG 834
137 LLeuAlaIleArgProLeuAlaLeuArgAlaValLeuGlyLeuGlyLeu 153
835 GTTCGCGCGCGCGCTCCCGCTGCTCATCTCCGACACTGTCGGCATC 884
154 .ProAlaAlaAsp....TrpGlyAlaValGlyArgTyrSer..ArgAs 167
885 CCGCGCGAGAGCACACCTGTGTCGCGCCACCGCGTGGCTCTTGAGGA 934
167 PValGlyArgThrLeuAspArgGlyAlaSerAlaGluAspMetArgArg 184
935 GCGCGGACACACGCTC.....GCGCGCGCGCACCGGTACGACGCG 975
184 LHisAlaAlaIleAlaGluPheAlaAspTyrValGlyAlaGlyAlaLeuAla 200
976 CCGAGGCGCGCTCCAGAGTTCAACCGCTACCTCGCGAGAGTGGAC 1025
201 ArgArgArgArgGluGlyGlyGluAspLeuLeuAlaLeuMetLeuAspAl 217
1026 CGCGCGCGCGCGCGACCGCGAGCATCTGCTACCTCTGCTCGCGCGC 1075
217 AHisAspArgGlyLeu..MetSerArgAsnGluIleValSerThrValY 233
1076 CCGGAGACACCGGATCACCGCTCACTGCGAGCGCATCTGCGACCTGCG 1125
233 AlThrPheIlePheThrGlyHisGluThrValAlaSerGlnValGlyAsn 249
1126 TCATCTGTCTACCGCGCGCGACGACACACCAACTGCTCGCGCAGG 1175
250 AlaValLeuSerLeuLeuAlaHisProAspGlnLeuAspLeuLeuArg 266
1176 GCGGTCTCACTGCGCGCGCGCGCTGCTGACAGTGGCGCAC 1225
266 GArgProAspLeuLeuAlaGlnAlaValGluGlyCysLeuArgTyrAsp 283
1226 CACACCGGATCGACACCGCGCGCGTGAAGATGATGCGGTACGAC 1275
283 roSerValGlnSerAsnThrArgGlnLeuAspValAlaGluLeuArg 299
1276 CGCCCGTGCAGCGGTGAGCGGTGGCGGTACAGACATCGCGCTCGGC 1325
300 GlyArgArgLeuAlaArgAspAspValAlaValAlaLeuAlaIleAla 316
1326 GACACGACATCCCGCGCGCGAGCGGTGTGTCGCTGTCGCGTCCGCG 1375
316 AAsnArgAspProAlaArgTyrArgArgArgProAspAspPheAspIleGlu 333
1376 GAACCGGACCGCGCGCTCCGCGCTCCGACGTCTGACGAGTCCGAC 1425
333 rGaAspProValProSerMetSerPheGlyAlaGlyMetArgTyrCysLeu 349
1426 GCGCGCGCGAAGGAGGTGGCTTCCGCTCGGAATCCACTGTGCTC 1475

```

```

350 GlySerTyrLeuAlaArgThrGlnLeuArgAlaAlaValAlaLeuAl 366
1476 GCGCGACCTGGCGCGCGCGAGCGCGAATCGCTCAAGGCGCTG... 1523
366 AArgLeuProGlyLeu...ArgLeuGlyCysAlaSerAspAlaLeuAlat 382
1524 ....CTGAGCGGATCCCGCGCTCGCGGAGCGCGCCACGAGTGCAGT 1569
382 yGlnProAlaGlyThrMetPheArgGly.....LeuAlaSerLeu 394
1570 ACGCCGACGACATGCTTCCAGCGCGCGAGCGCGCTCTCTCGACTG 1619
395 ProIleAlaPheThrPro 400
1620 CCGGAGCCACGTGCGCC 1637

seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:US-09-428-517-1
seq_documentation_block:
; Sequence 1, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Bectach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/428,517
; EARLIER FILING DATE: 1999-10-28
; EARLIER APPLICATION NUMBER: 60/120,254
; EARLIER FILING DATE: 1999-02-16
; EARLIER APPLICATION NUMBER: 60/106,100
; EARLIER FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA
US-09-428-517-1

alignment_scores:
Quality: 472.50 Length: 423
Ratio: 1.898 Gaps: 12
Percent Similarity: 58.865 Percent Identity: 34.043

alignment_block:
US-09-724-797-36 x US-09-428-517-1 ..
Align seg 1/1 to: US-09-428-517-1 from: 1 to: 50937
4 AspAlaValThrAlaPheAspProThrAspAlaAspValAlaArgAspR 20
46524 GACCGGTACCGCGCTAC...CGGTACAGCTCCGCGACGCGCTGACGT 46570
20 oTyrProSerTyrHisTrpLeuLeuArgHisAspProValHisArg.... 35
46571 CGACCCGACATACGCGCGACGCTCCGCGCGACGACGCGCTCCAGGTCG 46620
36 .....GlyAlaHisArgValIleTrpTyrValSerAlaPheAlaAsp 48
46621 GCGTCCCTTCGCGGAGCGAGCGAGCGCGTGGCTGTCACCGCATGTCGAC 46670
49 ValArgAlaValLeuGlyAspGluArgPheAlaAlaArgThrGlyTLeuArg 65
46671 GCGCGTATCGTTCGGGAGTACGCGCGTTCGACGACGCGCGCG.... 46715
65 gPheTrpThrAspLeuValGlyProGlyLeuLeu...AlaGluIleValG 81

```

```
46716 .....ACCGATCCCGCACCCCGGAGTGTTCACCAACCGCGCGAGC 46758
      ||||| : : : ||| : : : : :
81 LysPillelleuPheglnAspGluProAspHisGlyArgLeuArgGly 97
      ||| : : : ||| ||||| ||||| |||||
46759 CGGAGGCGCTCGCCAGAGACCGCGGACCAACCGCGTGGCGGG 46808
      : : : : : : : : : : : : : : :
98 ValValGlyProAlaPheSerProSerAlaLeuArgArgLeuGluProVa 114
      : : : ||||| : : : : : : : : :
46809 CTGGTGGGCAAGGCTTTCACGAGACCGCGGTGGAGAGATGGCGCCCG 46858
      : : : : : : : : : : : : : : :
114 LlleAlaGlyThrValAspAspLeuLeuArgProAlaLeuAlaArgGlyA 131
      : : : : : ||||| : : : : :
46859 TGTCCGCTCCCTCGTCACTCCCTGCTCGACAGCATGTGGCGCAGGTT 46908
      : : : : : : : : : : :
131 La...MetAspValValAspGluLeuAlaIatyrProLeuAlaLeuArgLa 146
      : : : : : ||||| : : : : :
46909 CACCGCGGACGCTGTGAGATTCTCGCGCTTCCCTCCCGCGGCTC 46958
      : : : : : : : : : : :
147 ValLeuGlyLeuLeuGlyLeuProAlaAlaAspTyrGlyAlaValGlyAr 163
      : : : ||||| : : : : :
46959 ATTCGCGACGTGCTCGCGCTGCGCTTGAGAGAC..... 46991
      : : : : : : : : :
163 gTTPSerArgAspValGlyArgThrLeuAsp..... 173
      ||||| : : : ||||| : : :
46992 .....CGGACCTGTTCGAGACTTTCACAGCCCATGCTCTCTCGA 47034
      : : : : : : : : :
174 ..ArgGlyAlaSerAlaGluAspMetArgArgGlyHisAlaAlaLea 189
      ||| : : : ||||| : : :
47035 CCGCGCTCACCGCGCGGAGATACAGCG.....GTCCAGCAG 47072
      : : : : : : : : :
190 GluPheAlaAspTyrValGluArgAlaLeuAlaArgArgArgGluG 206
      : : : ||| : : : : :
47073 GACTTCATGGTCTACATGAGCGCTGCTGCCAGCCCGCGAGCGCCC 47122
      : : : : : : : : :
206 yGlyGluAspLeuLeu..AlaLeuMetLeuAspAlaHisAspArgGlyL 222
      ||||| ||||| ||||| : : :
47123 CACCGAGAGCTGTGCGCGCTCGCCCTCGCCACCGCAACGACGAGC 47172
      : : : : : : : : :
222 eumetSerArgAsnGluIleValSerThrValValThrPheLeuPheThr 238
      : : : : : ||||| : : :
47173 ACCGACCAAGGCGGAGATGTCACATGAGGAGGAGGCTGCTCATGCGG 47222
      : : : : : : : : :
239 GlyHisGluThrValAlaSerGlnValGlyAsnAlaValLeuSerLeu 255
      ||||| ||||| : : : |||||
47223 GGCAGAGAGCGGTGCGTCAACGATCACCACTGCTCCACTCTGCT 47272
      : : : : : : : : :
255 uAlaHisProAspGlnLeuAspLeuLeuArgArgArgProAspLeuAla 272
      : : : : : : : : :
47273 GACCGAGCGCAAGCGCTACAGTCTGCTGCCGACCGCGCTCGTGC 47322
      : : : : : : : : :
272 laGlnAlaValGluGluCysLeuArgTyrAspProSerValGlnSerAsn 288
      ||||| ||||| ||||| : : :
47323 CCGCGCGGTGGAGAGATGCTGCGGTACACACCGCTGTGTCGCCCGC 47372
      : : : : : : : : :
289 Thr.....ArgGlnLeuAspValAspValGluLeuArgGlyArgGle 303
      : : : ||| : : : ||||| : : :
47373 AGCTTCCTCGCGCTGCGCACCGAGCTGAGCTGACACCGCTGACCTG 47422
      : : : : : : : : :
303 uArgArgAspAspValValValValLeuAlaGlyAlaAlaAsnArgAspP 320
      : : : : : ||||| : : : |||||
47423 GCGGCGCGGGAGACCTGCGTGTCTCATTCCGTCGCGCAACGCGAGC 47472
      : : : : : : : : :
320 roArgArgGlyTyrAspArgProAspAspPheAspIleGluArgAspProVal 336
      : : : : : : : : :
47473 AGGAGGCTTCGACACGCGGAGACTGCTGACCTTCACCGTGAAGCGAAC 47522
      : : : : : : : : :
337 ProSerMetSerPheGlyAlaGlyMetArgTyrCysLeuGlySerTyrLe 353
      ||| : : : ||||| ||| : : : ||||| : : :
47523 CCGGACATACGCTTCGCGGACGAGCGACCACTGATCGCGCCCACT 47572
      : : : : : : : : :
353 uAlaArgThrGlnLeuArgAlaAlaValAlaLeuAla..ArgLeuP 369
      : : : ||| : : : : : ||||| : : : |||||
```

```
47573 GAGCGACTGGAAGCTCCAGAGAGCCCTGTCCGCCCTGTCCGCGCTTCC 47622
369 roGlyLeuArgLeuGlyCysAlaSerAspAlaLeuAlaIatyrGlnProArg 385
      ||| ||| : : : : : : : : :
47623 CCACCTGTGATGTGGCCAGCGCGGTGCGGGAGTGAAGGAGCAGGCG 47672
      : : : : : : : : :
386 ThrMetPheArgGlyLeuAlaSer.....LeuProI 396
      : : : : : ||||| : : :
47673 ATGCTGATCCCGGAGATGGAAGCCAGATGCTGCTCGGTGAGAGCGCGGC 47722
      : : : : : : : : :
396 eAlaPheThrProGlyGly 402
      ||| : : : ||||| : : :
47723 CGCCGCGCGCGCGCGCGG 47741
```

seq\_name: /cgn2\_6/prodata/1/lna/Backfile1.seq:5212296-17

seq\_documentation\_block:

```
: Patent No. 5212296
: APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LIETO, KENNETH
: J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSEK, JAMES A.
: TEPPERMAN, JAMES M.
: TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
: CYTOCHROMES
: NUMBER OF SEQUENCES: 19
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/569,781
: FILING DATE: 23-AUG-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 464,499
: FILING DATE: 12-JAN-1990
: APPLICATION NUMBER: 405,605
: FILING DATE: 11-SEP-1989
: SEQ ID NO:17:
: LENGTH: 1212
5212296-17
```

alignment\_scores:

Quality:	454.00	Length:	389
Ratio:	1.924	Gaps:	10
Percent Similarity:	60.668	Percent Identity:	34.447

alignment\_block:

US-09-724-797-36 x 5212296-17 ..

Align seg 1/1 to: 5212296-17 from: 1 to: 1212

```
19 AspProTyrProSerTyrHisThrPheLeuArgHisAspProValHisAr 35
      ||||| ||||| : : : ||||| : : :
76 GACCCCGCGCGCGAGCTCACGAGCGCGCGCGCGACCGAACCGGTCAACCG 125
      : : : : : : : : :
35 gGly.....AlaHisArgValTyrTyrValSerArgPheAla 48
      : : : : : : : : :
126 GGCACACCTGTGGAGCGGCTCTCTCTGTGCTGTGAGCCGCCATCAG 175
      : : : : : : : : :
48 spValArgAlaValLeuGlyAspGluArgPheAlaArgThrGlyLeuArg 64
      ||||| ||||| ||||| : : : |||||
176 AGCTCCGCGCGGCTCTCTCTGAGCCCGCGCTTCAGCGCCGACGCGCACCG 225
      : : : : : : : : :
65 ArgPheThrPheAspLeuValGlyProGlyLeuLeuAlaGluIleValG 81
      : : : : : : : : :
226 ACCGCTTCCTCTCTGACCGCGCGGCG.....CGGAGATCATCGG 269
      : : : : : : : : :
81 Y.....AspIlelleuPheGlnAspGluProAspHisGlyArgLeu 96
      : : : : : : : : :
270 CACCAACCGGACCTTCTCGGCAATGAGACCGCGGACCGCGCCGCTGC 319
      : : : : : : : : :
96 rgGlyValValGlyProAlaPheSerProSerAlaLeuArgArgLeuGlu 112
      ||| : : : ||||| : : : : :
320 GCGGATGCTCACCGCGGACTTCACTGTCGAAGAAGTTCAGAGCATGCGC 369
      : : : : : : : : :
113 ProValIleAlaGlyThrValAspAspLeuLeu..ArgProAlaLeuAl 128
      ||| : : : ||||| : : : |||||
```

```
370 CCCGAGGTGCAGCGCTCCGCGACGACCTGTCGACCGGATGACCACCGG 419
128 aargllylmetaspvalvalaspglutlualatyprtleuallaleua 145
420 AGCGACCTCCCGCCGCTGGTACCGAGTTGGCGTGGCGGCTGCCGCTCC 469
145 rgalavalleuylleuenglyleuproalalalaspptrglyalaval 161
470 TGGTATGCTGCTGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 519
162 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 178
520 CAGGAGCGCGCGCGGCTCTG.....CTCACCTCGCGCTCCAGTCC 560
178 agluspmetatargarglylhisalalalealagluphealasptryv 195
561 CGAGGAGATGCGG.....GCCGCCAGGACGAGTGGTGAGTAC 601
195 aiglurgalaleualatargatargatgluglyglylunspheuleu 211
602 TGGCCCGG...CTGCGCGGACCAAGCGGAGCGCGCGCGCGCGCGCATC 648
212 AlaleuMetleuaspaahisaspargglyleuMetSerArgasnlu1 228
649 ATCAGCGCGCTGCTGCC.....CGCGGAGCTGAGACACCAAT 692
228 eValSerthrvalvalthrPheIlePheThrlylhisgluthrvalalas 245
693 CGCCACCATGCGGACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 742
245 ercInValglYasnAlaValleuSerleuAlaHisProaspInleu 261
743 ACATGACCGGCTCTCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 792
262 AspleuLeuarghrargproaspLeuAlaValglucy 278
793 GCCCGGCTGCGCGGACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 842
278 sleuargtYrAspProserValGlnSeranThrargInleuAspVal. 294
843 GCTGCGTACCTGACATCGTCGACAAAGCGCTTCCCGGATGCCACCG 892
295 .AspValGluLeuArglylArgArgLeuArgArgAspAspValValVal 310
893 AGGACGCTCATCGCGCGCGACATCGCGCGCGGAGGCGCTGCTG 942
311 ValLeuAlaGlyAlaAlaAsnArgAspProArgArgTyrAspArgProAs 327
943 TGCATGATCAAGCTCCGCAACCGGAGCGCGAGTGTCCCGCGGACGA 992
327 PAspPheAspIleGluArgAspProValProSerMetSerPheGlyAlaG 344
993 CGACCTGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCTTGGCTGCG 1042
344 lYmetArgTyrCysLeuGlySerTyrLeuAlaArgThrInleuAlaGala 360
1043 GCGTCACAGTGGCTGGGACAGCGCTTGGCGCGGAGGAGCTTCAGATC 1092
361 AlAlaVal...AlaAlaLeuAlaArgLeuProGlyLeuArgLeuGlyCysAl 376
1093 GCCATCGAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1142
376 aSerAspAlaLeuAlaTyrGlnProArgThrMetPheArgGlyLeuAlas 393
1143 CCAGGAGAGATCCCGTTCCGCGCGCGACATGGCATCTACGGGGTCCACT 1192
393 erleuproIleAlaPhe 398
1193 CGTCTCGATCGCTGG 1209
seq_name: /cgn2_6/ptodata/1/lna/5B_COMB.seq:us-09-103-840A-2
seq_documentation_block:
```

```
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
```

```
alignment_scores:
Quality: 446.00 Length: 414
Ratio: 1.890 Gaps: 11
Percent Similarity: 57.005 Percent Identity: 31.401
```

```
alignment_block:
US-09-724-797-36 x US-09-103-840A-2/rev ..
```

```
Align seg 1/1 to reverse of: US-09-103-840A-2 from: 1 to: 4403765
```

```
6 valthrAlaPheaspProthrAspAlaaspValalargargaspProtyrP 22
2541149GTCACCGACTTGACCCGCTTGATGCCGCGATGCGGTGATCTTACCC 25411100
22 cSerTyrHisThrLeuLeuArgHisaspProvalHisArgGlyAlaHisA 39
2541099GATATATGAGAGCTGTGCGCGGCGGAGCGGCTGCTGCTGCTGCTG 25410500
39 rg...ValThrPyrValSerArgPheAlaaspValArg.....AlaVal 52
2541049GCGAGCTGTACATTTTGAGCCGATACGCCAGCTCCGCGAGCGCTGCG 25410000
53 LeuGlyAspGluArgPheAlaArgThrGlyLe.....ArgArgPheTr 67
2540999AACCAACGACGCTGTCCAGCGCGCGCGGCTGCTGCTGCTGCTGCTG 25409500
67 pThrAspLeuValGlyProGlyLeuLeuAlaGluIleValGlyaspIleI 84
2540949GCTGCGCTTCTGCGGACG..... 2540931
84 leuLeuPheGlnaspGluProaspHisGlyArgLeuArgGlyValValGly 100
2540930.....TCCGATCCGCGCGCGCGCACACCGGATGCCGACGACACTG 2540889
101 ProAlaPheSerProSerAlaLeuArgArgLeuGluProValIleAlaG 117
2540888CCAGGATATGCGCGCGCTTGGAGAGCTGCGCGCGCGATGATGTCAC 2540839
117 yThrValAspAspLeuLeuArgProAlaLeuAlaArgGlyAlaMetAspY 134
2540838GCTTCCCGCAAGCTGTGCGGCGGCTTACTGTACCCAGACGCGCGGAG 2540789
134 alValaspGluLeuAlaTyrProLeuAlaLeuArgAlaValleuGlyLeu 150
2540788TGTCTTCACCGCGCGCGCGATGCCGCTTACACAGTGTG 2540739
151 leuGlyLeuProAlaAlaaspTrrgLyAlaValGlyArgTrrSerArgAs 167
```





[illegible]

```

Alignment_scores:      Quality:   427.00      Length:    393
                       Ratio:     1.825       Gaps:        12
Percent Similarity:    59.542      Percent Identity: 34.097

alignment_block:
US-09-724-797-36 x 5212296-8  ..

Align seg 1/1 to: 5212296-8 from: 1 to: 1998

19 AsprProTyrProSerTyrHisTrpLeuLysArgHisAspProValHisAr 35
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
270 GACCCCGGCGGCCAGTCACGGAGCGGGCGCCGACGACCGGTCAACCCG 319
35 ggly.....AlahisArgValTrpTyrValSerArgPheMetIaa 48
|::: |::: |::: |::: |::: |::: |::: |::: |::: |:::
320 GCCCACCTCTGGGAGCGCTCTCTGCTGCTGTATACGGGCATCAGG 369
48 sPvalArgAlaValLeuLysArgLysArgPheAlaArgThrGlyLearg 64
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
370 AGCTCGGCGCGGTCTCGGGCAGCCCGCGCTTACAGCGCGACGCCAACGC 419
65 ArgPheTrpTrpHisPLeuValGlyProGlyLeuLeuIacLIleValGI 81
::: |::: |::: |::: |::: |::: |::: |::: |::: |:::
420 ACCGGCTTCCCTTCTCTGACCGCGCGGCGC.....CGCGAGATCATCGG 463
81 y.....AspIleIleLeuPheGlnAspGluProAspRnIGLyArgLeu 96
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
464 CACCACCCGACCTTCTCGCGCATGGAGACGACCGGACCGCCGACTGC 513
514 GCGGAGTCTCAGCGCGCATCTTCAATGTCAGAGAAGTCCAGGCGATCGC 563
96 rglYValValGlyProLarPheSerProSerAlaLeuArgArgLeuGI 112
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
514 GCGGAGTCTCAGCGCGCATCTTCAATGTCAGAGAAGTCCAGGCGATCGC 563
113 ProValIlealaGlyThrValAspAspLeu...ArgProalaLeuAl 128
|| |::: |::: |::: |::: |::: |::: |::: |::: |::: |:::
564 CCGGAGGTGCAGCGCTCTCGCGACGACCTGGTCCAGCGATGACACACCG 613
128 argGLyIlaMetasrPvalIlaAspGlyLeuAlaTyProleuAlaleuA 145
||||| |::: |::: |::: |::: |::: |::: |::: |::: |:::
614 ACGCACCTCGCGCGACCTGGTCCACCGAACTTGCGCTCGCTCGCTCCC 663
145 rgalValalLeuGIyreuleuGIyreuproAlaIlaAspTrpGlyalVal 161
||||| |::: |::: |::: |::: |::: |::: |::: |::: |:::
664 TCGTATGTGCGCTGCTGGCGGCGCTTCAAGAGACAAGCGCTTCTC 713
162 GlYArgTrpSerArgArgPValGIYAArgThyleuAspArgGlyAlaSerAl 178
::: |::: |::: |::: |::: |::: |::: |::: |::: |:::
714 CAGGAGCGCAGCGCGGTCTG.....CTCACCTCGCGGTCTCACTCC 754
178 agLysrMetArgdArgGlyHisAlaIlaIlealGIyPheIlaSrrTyT 195
||||| |::: |::: |::: |::: |::: |::: |::: |::: |:::
755 CGAGGAAGTCGG.....GCGCGCCAGACGAGCTGTGGAGTAAC 795
195 aGlArgAlaLeuAlaIlaArgArgArgArgGlyGIYGIYAspLeuLeu 211
::: || | | | | | | | | | | | | | | | | | | | | | |
796 TCGCCCCG..CTCGCGCGACCAAGAGCGAGCGCGCGCGAGACCGATC 842
212 AlaleuMetLeuAspAlaHisAspArgGlyLeuMetSerArgArgGngLI 228
|| | | | | | | | | | | | | | | | | | | | | | |
843 ATCACACCCCTGTGTGTC.....CGCGCGAGCTCGACGACACCCAGAT 866
228 eValSerThrValValThrPheIlePheThrGlyHis..GIuThrAla 244
|::: |::: |::: |::: |::: |::: |::: |::: |::: |:::
887 CGCCACCATGGGACGCTGTGGTGTGTGTCGCGCGCGGACCTCGACG 936
244 IasergLIvalGlysnAlaValalLeuSerLeuLeuAlaHisArgDgIn 260
::: | | | | | | | | | | | | | | | | | | | | | |
937 ACACCCAGATGCGCCACATGGGATG..CTGTGCGGCAACCCGACCAA 983
261 LeuAspLeuLeuArgArgArgProAspLeuLeuIacLIleValalGIuGI 277
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

```



```

312 euAlaGlyAlaAlaAsnArgAspProArgTyrAspArgProAsp 328
   :: :: ||||| ||||| :: :: :: |||||
956 TCAACTCATAGCCACCGGAGCGACGGGTGACGAGACCGGAGCC 1005
   :: :: ||||| ||||| :: :: :: |||||
339 PheAspIleGluArgAspProValProSerMetSerPheGlyAlaGly 345
   :: :: ||||| ||||| :: :: :: |||||
1006 CTGCAATCAACCGCTCGCGCGGACACCTGCTTCGCGCTCGGCGT 1055
   :: :: ||||| ||||| :: :: :: |||||
345 TArgTyrGlySerGlySerTyrLeuAlaIleGlyThrGlnLeuArgAlaVal 362
   :: :: ||||| ||||| :: :: :: |||||
1056 GACCAAGTGGCTGGGCGGAGACCTCGCGGCTGAGCTGAGGTCTCC 1105
   :: :: ||||| ||||| :: :: :: |||||
362 AlaAlaAlaLeuAla...ArgLeuProGlyLeuArgLeuGlyCysAlaSer 377
   :: :: ||||| ||||| :: :: :: |||||
1106 TCAACGCCCTCATGGACCGGCTCCGAGACGCTGGCGCTGCCGCTC 1155
   :: :: ||||| ||||| :: :: :: |||||
378 AspAlaLeuAlaIleTyrGlnProArgTyrMetPheArgGlyLeuAlaSer 394
   :: :: ||||| ||||| :: :: :: |||||
1156 GAGCAGTTGGTGGCTGGCGCGGTACGACATCCAGGCGCTCAACGACT 1205
   :: :: ||||| ||||| :: :: :: |||||
394 uProIleAlaPhe 398
   ||||| :: :: :: |||||
1206 CCGGCTCACTGG 1218

seq_name: /cgn2_6/prodata/1/ina/backfiles1.seq:5212296-5
seq_documentation_block:
: Patent No. 5212296
: APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LETO, KENNETH
: J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSER, JAMES A.
: TERPMAN, JAMES M.
: TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
: CYTOCHROMES
: NUMBER OF SEQUENCES: 19
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/569,781
: FILING DATE: 23-AUG-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 464,499
: FILING DATE: 12-JAN-1990
: APPLICATION NUMBER: 405,605
: FILING DATE: 11-SEP-1989
: SEQ ID NO: 5
: LENGTH: 1879
5212296-5

alignment_scores:
: Quality: 417.50 Length: 421
: Ratio: 1.625 Gaps: 13
Percent Similarity: 61.045 Percent Identity: 30.166

alignment_block:
US-09-724-797-36 x 5212296-5 ..
Align seg 1/1 to: 5212296-5 from: 1 to: 1879

2 LeuValAspAlaValThrAlaPheAspProThrAspAlaVal..... 16
: :: :: ||||| ||||| :: :: :: |||||
128 ATGACGATACCGCCAGCAGCGCCAGACGACGACCGCGCTTCCC 177
   :: :: ||||| ||||| :: :: :: |||||
17 .....ArgArgAspProTyr.....ProSerTyrHisTrpLeuLeu 29
   ||||| ||||| :: :: :: |||||
178 GAGCAACCGGAGCTGCTCCCTACGATTAACCGGAGCGCTACGCCAGCTCC 227
   ||||| ||||| :: :: :: |||||
29 rghIs.....AspProValHisArg.....GlyAlaHisArg 39
   ||||| ||||| :: :: :: |||||
228 GGGACACCCCGCGCCCTGACCGGCTGACGCTTACGACGCGCGCTCAG 277
   :: :: ||||| ||||| :: :: :: |||||
40 ValTrpTyrValSerArgPheAlaAspValArgAlaValLeuGlyAsp 56
   :: :: ||||| ||||| :: :: :: |||||
278 GCGTGGGTGTGACCAAGCAGACGCGCGCGCAACTGCTCGGCGACCC 327

```

```

56 uArgPheAla.....ArgThrGlyIleArg.....A 65
   :: :: ||||| ||||| :: :: :: |||||
328 CCGGCTGTCTCAACCGGAGCGACGACACTTCCCGCCAGCTACCGC 377
   :: :: ||||| ||||| :: :: :: |||||
65 rPheTrpThrAspLeuValGlyProGlyLeuLeuAlaIleValGly 81
   ||||| :: :: :: |||||
378 GCTTCGAGCGCGTCCGGGAGAGCCG..... 403
   :: :: ||||| ||||| :: :: :: |||||
82 AspIleIleLeuPheGlnAspGluProAspHisGlyArgLeuArgGly 98
   :: :: ||||| ||||| :: :: :: |||||
404 CAGGGCTTCATCGGCTGGACCCCGGACAGCGACCGCGGCGGTAT 453
   :: :: ||||| ||||| :: :: :: |||||
98 ValGlyProAlaPheSerProSerAlaLeuArgArgLeuProVal 115
   ||||| :: :: :: |||||
454 GACGATCGCGCATGTTCAACCGGATCAAGCGGATCAAGCGGATCGCC 503
   :: :: ||||| ||||| :: :: :: |||||
504 TCGAGGAGGTGTGTCACGGCTTCTTCGACGAGATGCTGGCGCGCGC 553
   :: :: ||||| ||||| :: :: :: |||||
131 AlaMetAspValValAspGluLeuAlaTyrProLeuAlaLeuArgAla 147
   :: :: ||||| ||||| :: :: :: |||||
554 ACCGCGACCTGGTCACTGCTGCGCTGCGCTGCTGCTGCTGCTGCT 603
   :: :: ||||| ||||| :: :: :: |||||
147 IleuGlyLeuLeuGlyLeuProAlaAlaAspTrpGlyAlaValGlyArg 164
   :: :: ||||| ||||| :: :: :: |||||
604 CTGCGGACTCTCTCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCT 641
   :: :: ||||| ||||| :: :: :: |||||
164 rPseAlaGlyAspValGlyArgThrLeuAspArgGlyAlaSerAlaGlu 180
   :: :: ||||| ||||| :: :: :: |||||
642 TCTTCAGGACGCGAGCAAGGCGCTG.....GTGCGTCCAGCGAC 682
   :: :: ||||| ||||| :: :: :: |||||
181 MetArgArgGlyHisAlaAlaIleAlaGluPheAlaAspTyrValGlu 197
   :: :: ||||| ||||| :: :: :: |||||
683 GCGCAGAGCGGCTGACCGCGGAGACGCTGCGGCTTACCTGAGCG 732
   :: :: ||||| ||||| :: :: :: |||||
197 GAlaLeuAlaArgArgArgArgGlyGlyGlyAlaSerPheLeu...Ala 213
   :: :: ||||| ||||| :: :: :: |||||
733 CCGTATCAACCGGATCCAGACGCGGAGCGGCGGCTGCTGCGGCTG 782
   :: :: ||||| ||||| :: :: :: |||||
213 eumetLeuAspAlaHisAspArgGlyLeuMetSerArgAsnGluIleVal 229
   :: :: ||||| ||||| :: :: :: |||||
783 TGTTCGCGGACGACCTGCGGAGCGGAGATGACGCTGAGGAGATGAT 832
   :: :: ||||| ||||| :: :: :: |||||
230 SerThrValValThrPheIlePheThrGlyHisGluThrValAlaSer 246
   :: :: ||||| ||||| :: :: :: |||||
833 TCCACCGGAGTGTCTCTCATGCGCGGACGAGACGCGGCTGAT 882
   :: :: ||||| ||||| :: :: :: |||||
246 nValGlyAsnAlaValLeuSerLeuAlaHisProAspGluLeuAsp 263
   :: :: ||||| ||||| :: :: :: |||||
883 GACCTCCCTGAGCTGATCAACCTGCTGAGCACACCCCGAGCATACGCG 932
   :: :: ||||| ||||| :: :: :: |||||
263 euleuArgArgArgProAspLeuLeuAlaGlnAlaValGluGlyCysLeu 279
   ||||| :: :: :: |||||
933 CCTGCGCGCGGACCGCGGAGCTGCGCGCGCGGTGAGGAGAACTGCT 982
   :: :: ||||| ||||| :: :: :: |||||
280 ArgTyrAspProSerValGln...SerAsnThrArgGluLeuAspVal 295
   ||||| :: :: :: |||||
983 CGCTACCTCGCATCGCGGACATCGCGGCGCGCGCTGCCACGCGGGA 1032
   :: :: ||||| ||||| :: :: :: |||||
295 pValGluLeuArgGlyArgArgArgArgArgAspAspValValVal 312
   :: :: ||||| ||||| :: :: :: |||||
1033 CATCGAGGTGAGGAGGACCTCATCGGCGGCGGAGGCGGTGATGCTG 1082
   :: :: ||||| ||||| :: :: :: |||||
312 euAlaGlyAlaAlaAsnArgAspProArgArgTyrArgArgProAspArg 328
   :: :: ||||| ||||| :: :: :: |||||
1083 TCAACTGATAGCCACCGGAGCGGACGCTGTACGAGACCGCGCGCTC 1132
   :: :: ||||| ||||| :: :: :: |||||
329 PheAspIleGluArgAspProValProSerMetSerPheGlyAlaGly 345
   :: :: ||||| ||||| :: :: :: |||||
1133 CTGACATCAACCGGCTCGCGCGGCGGACCAACCTGCTGCGGCTGCGCT 1182

```

```

seq_name: /cgn2_6/plodata/1/ina_5A.COMB.seq:US-08-102-863-10
seq_documentation_block:
; Sequence 10, Application US/08102863
; Patent No. 5466590
GENERAL INFORMATION:
APPLICANT: SARISIANI, SIMA
TITLE OF INVENTION: CONSTITUTIVE
TITLE OF INVENTION: EXPRESSION OF P450CYP
TITLE OF INVENTION: AND FERREDOXIN-SOY IN
TITLE OF INVENTION: STREPOMYCES
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS
ADDRESSEE: AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: USA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Version 1.0,
SOFTWARE: Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/102,863
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/807,001
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: GALEBOIS, R. THOMAS
REGISTRATION NUMBER: 32,692
REFERENCE/DOCKET NUMBER: CR-9000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-7342
TELEFAX: 302-892-7949
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1735 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-102-863-10

alignment_scores:
Quality: 410.00 Length: 428
Ratio: 1.760 Gaps: 10
Percent Similarity: 54.439 Percent Identity: 29.439

Alignment_block:
US-09-724-797-36 x US-08-102-863-10 ..

```

```

Align seq /1 to: US-08-102-863-10 from 1 to: 1735

10 ASPProthThAspAlaAspValAlaArgAlaAspProTyrProSer..... 23
    |||||
    |||||
223 GACCCACCC.....TTCGGGCCCCCGGCGAGCTCTT 254
225 CCGGACGAGACCGGGGGTGGCCCTTACACACGCGCCGGGTGACGACCGC 304
    |||||
    |||||
33 aHisArgLy.....AlaHisArg 39
    :|:|:|
    :|:|:|
305 TGGCGAGGGCCCGCGGTAGCCGGGTACCCCTTTCAGACGACGCCG 354
40 ValTrrTyrValSerArgPheAlaAspValAlaValLeuGlySp.. 55
    |||||
    |||||
355 GTCTGGCGGCTTACACGGGACAGCCCTGGGCCCTGAGCTACTGGCGGACCC 404
56 .....GluA 57
405 GCGGCTCTCCACCGACCGACCGACCCGAGCTTCCCGTCCGCGGAC 454
57 rrpheaIaArgThrGlyIleArgArgPheThrPheAspLeuValGlyPro 73
    |||||
    |||||
455 GGTTCGGCGCGCGGACGGCGCGCGCTC..... 483
74 GlyLeuLeuAlaGluIleValGlyAspIleIleLeuPheGlnAspGluP 90
    |||||
    |||||
484 .....GCCTCTCTGCGCTCAGCACGCC 506
90 oAspHisGlyArgLeuArgGlyValValGlyProAlaPheSerProSerA 107
    |||||
    |||||
507 CGACACACACACCGACCGACGAGATGCTCATCCCGACTTCTCGGTAGC 556
107 IaleuArgArgLeuGluProValIleAlaGlyThrValAspAspLeu 123
    :|:|
    :|:|
557 GGAATCGCGGCTCCGCCCGCTATCCAGAGACCGTGGACCGGCTCTC 606
124 ArgProAlaLeuAlaArgLy...AlaMetAspValValAspGluLeuAl 139
    |||||
    |||||
607 GACGCGATGGACGACACAGGAGGCCCGCGGCAACTGGTGAAGCCGTTCCG 656
139 aTyrProLeuAlaLeuArgAlaValLeuGlyLeuGlyLeuProAlaA 156
    |||||
    |||||
657 CMTCCGCGTCCCGTGCATGTGATCTGTCTGCTCGCGGCTCCCTACG 706
156 IAspTrrPglValAlaValGlyArgTrrSerArgAspValGlyArgThrLe 172
    |||||
    |||||
707 CCGGACACCGG.....TTCCTGAGGAGACGCTGGACGAGATC 744
173 AspArgGlyAlaSerAlaGluAspMetArgArgGlyHisAlaIleAl 189
    |||||
    |||||
745 CTGGCGGCGCCGGGAGCGGACGATGTGAACAG.....GCCGCGA 785
189 aGluPheAlaAspTyrValGluArgAlaLeuAlaIaArgArgArgGluG 206
    |||||
    |||||
786 CGAAGCTGAGAGTAATCTGGGCGCGCTGATCGACCGGACAGAGGCGGAGC 835
206 IyGlyGlyAspLeuLeu...AlaLeuMetLeuAspAlaHisAspArgLy 221
    |||||
    |||||
836 CGGCTGACGGCTCTCTGACGAGGTATTCACCGGAGACACCGGAGCGA 885
222 LeuMetSerArgAsnGluIleValSerThrValValThrPheIlePheTh 238
    :|:|:|
    :|:|:|
886 CCGGTGACCGCGAAGAGCTGTGCTTGGCCCTTGCCTATCCCTGCATCCG 939
238 rGlyHisGlyThrValAlaSerGluValGlyAsnAlaValLeuSerLeu 255
    |||||
    |||||
936 CGGCGACGAGACGACGGGGAACATGATCTGGCTCGGACGTTACAGCTGC 985
255 euaIaHisProAspGluLeuAspLeuArgArgArgProAspLeu 271
    |||||
    |||||

```

```

986 TGAGCACCACCCGAAACAGCTGGCGGCGCTGGGGCCGCGGAGCAGCACC 1035
272 ALAAlaAlaValAlaGluCysLeuArgTyrAspProSerValGlnSerAs 288
1036 GCCGTGTGTTCGAGGAGCAGTGTCTCGGTTCTCTCATCGCGAGGGCCT 1085
288 ntrArgGlnLeuAspValAlaGluLeuArgGlyArgArgLeuArgA 305
1086 CCAGCGCGCTGGCAGCAGCAGATGAGTGCAGCGGGGCGACATCCGCA 1135
305 rgAspAspValAlaValAlaValAlaGlnAlaAlaAsnArgAspProArg 321
1136 AGGGGAGGGCGGTGTCTTCTTCACCTCGCTGATCAACCGCGACGCCGAC 1185
322 ArgTyrAspArgProAspAspAspIleGluArgAspProValProSe 338
1186 GTGTTCGCCCGGGCGGAGACATGACTGGAGCCGCCGCCGCATCA 1235
338 rMetSerPheGlyAlaGlyMetArgTyrCysLeuGlySerTyrLeuAla 355
1236 CCTGCGCTTGGCTTCGAGTCCAGCCAGTGTCTGGCGGACCACTGGCCC 1285
355 rgrHrGlnLeuArgAlaAlaValAlaAlaLeu...AlaArgLeuProGly 370
1286 GCCCGAGCTGACATCGCATCGGACCTCTTCGAGCGGCTTCCGGG 1335
371 LeuArgLeuGlyCysAlaSerAspAlaLeuAlaTyrGlnProArgThrMe 387
1336 CTCAGGCTCGCGTACCCGCGCAGAGATCCGTCAACGCGGGGAGCAC 1385
387 tPheArgGlyLeuAlaSerLeuProIleAlaPhe 398
1386 GATCCAGGGCTCTCTCAGACTGCCCGTGGCTGG 1419
seq_name: /cgn2_6/protoata/1/lna/PCTUS_COMB.seq:PCT-US92-10885-10

```

```

seq_documentation block:
; Sequence 10. Application PC/TUS9210885
; GENERAL INFORMATION:
; APPLICANT: SARIASLANI, SIMA
; TITLE OF INVENTION: CONSTITUTIVE
; TITLE OF INVENTION: EXPRESSION OF P450SOY
; TITLE OF INVENTION: AND FERREDOXIN-SOY IN
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS
; ADDRESS: AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: USA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch,
; MEDIUM TYPE: 1.0 MB
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh System, 6.0
; SOFTWARE: Microsoft Word, 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/10885
; FILING DATE: 19921216
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: GALLEGOS, R. THOMAS
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: CR-9000-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-7342
; TELEFAX: 302-892-7949
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1735 base pairs

```

```

; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US92-10885-10

alignment_scores:
  Quality: 410.00      Length: 428
  Ratio: 1.760        Gaps: 10
  Percent Similarity: 54.439  Percent Identity: 29.439

alignment block:
US-09-724-797-36 x PCT-US92-10885-10
Align seg 1/1 to: PCT-US92-10885-10 from: 1 to: 1735

10 AspProThrAspAlaAspValAlaArgAspProTyrProSer..... 23
223 GACCCACAC.....TCCCGCGCCCGCGGAGCTCCTT 254
24 .....TyrHisTyrLeuLeuArgHisAspProV 33
255 CCCGAGAGACCGGGGTGCCCCCTTACCCGCCCGCGGAGCAGCCG 304
33 alHisArgGly.....AlaHisArg 39
305 TCCGCGAGGCGCGCGCGGTGACCGGGTCACTTTCGAGGAGCGCCG 354
40 ValTyrPThrValSerArgPheAlaAspValAlaValLeuGlyAsp.. 55
355 GTCTGGGCGTACCGGGCAGCGCCCTGCGCTGCTACTGGCGGAGCC 404
56 .....GluA 57
405 GCGGCTCTCCACCGAGCCGAGCCAGCCGACTTCCCGTCCCGGCGAGC 454
57 rPheAlaArgThrGlyIleArgArgPheTyrPThrAspLeuValGlyPro 73
455 GGTTCGCGCGCGGAGCGCGCGCGCGCTC..... 483
74 GlyLeuLeuAlaGluIleValAlaGlyAspIleIleLeuPheGlnAspGluPr 90
484 .....GCTGTGCTCGCGGTGAGCAGACC 506
90 oAspHisGlyArgLeuArgGlyValAlaGlyProAlaPheSerProSerA 107
507 CGAGCACAACACCCAGCGCGAGTGTCTATCCGACCTTCTCGTGAAAC 556
107 LeuLeuArgArgLeuGluProValIleAlaGlyThrValAspAspLeu 123
557 GGATCGCGCGCTCGCGCGGTATCCAGAGACCGTGAACGCGCTCTC 606
124 ArgProAlaLeuAlaArgGly...AlaMetAspValAlaAspGluLeuAl 139
607 GACCGCATGAGGAGCAGACAGGGCCCCCGCGCACTGTGAGCGCGTCC 656
139 aTyrProLeuAlaLeuArgAlaValLeuGlyLeuGlyLeuProAlaA 156
657 CTGCGCGGTCCGCTCGATGTGTGTCTGTCTGCTGCGCGGCTCCAG 706
156 lAspTyrPThrGlyAlaValAlaGlyArgTyrSerArgAspValAlaGlyArgThrLeu 172
707 CCGACACACGCG.....TTCTTCGAGGAGACGCTCGCAGCAGCTC 744
173 AspArgGlyAlaSerAlaGlnAspMetArgArgGlyHisAlaAlaLeuAl 189
745 CTGCGCGCGCGGAGCGGAGCGAGTGTGAACAGG.....GCCGCGA 785
189 agluPheAlaAspTyrValGluArgAlaLeuAlaArgArgArgGluG 206
786 CGAAGTCTGAGGAGTACTGGCGCGCTGATCGAAGCGAAGGGCGGAGC 835

```







OM of: US-09-724-797-36 to: EST:\* out\_format: pfs

Date: Jun 21, 2002 5:40 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

# Command line parameters:

```
-MODE=frame-p2n.model -DEV=xlh
-O/cgn2.1/USPTO.spool/US09724797/runat-21062002.092538.18297/app-query.fasta.1.466
-DB=EST -OFMT=fastap -SUFFIX=1st -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELXT=6.000
-DELXT=7.000 -STARF=1 -MATRIX=blotsum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZ=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09724797.cgn1.1.6398
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLIPXY -WAIT -THREADS=1
```

# Search information block:

```
Query: US-09-724-797-36
Query length: 402
Database: EST:*
Database sequences: 13736207
Database length: -1841457050
Search time (sec): 1687.050000
```

# Score list:

Sequence	Strd	Orig	ZScore	Escore	Len	Document
gb_hlc:AK004861	+	203.00	313.09	3.4e-08	3004	AK004861 Mus musculus adult ma
gb_est1:AA886024	+	187.00	300.27	1.8e-07	768	AA886024 uc72b04.y1 Sugano mus
gb_est1:AI330226	+	184.00	295.36	3.3e-07	762	AI330226 u190e12.y1 Sugano mus
gb_hlc:BC001776	+	182.50	280.94	2.1e-06	2492	BC001776 Homo sapiens, similar
gb_hlc:BC020267	+	182.50	280.94	2.1e-06	2492	BC020267 Homo sapiens, clone 1
gb_est1:BI246899	+	182.00	292.62	4.7e-07	717	BI246899 602959455f1 NCI, CGAP-I
gb_est1:BG204536	+	179.50	287.54	9.1e-07	789	BG204536 RST73942 Athersys RAGE
gb_est1:BG206832	+	179.00	287.15	9.5e-07	755	BG206832 RST76133 Athersys RAGE
gb_est1:BG182832	+	179.00	287.02	9.7e-07	765	BG182832 RST1693 Athersys RAGE
gb_est1:AI331314	+	177.50	283.91	1.4e-06	814	AI331314 m187b03.y1 Soares mus
gb_hlc:BI619542	+	176.50	282.07	1.8e-06	828	BI619542 RST15025 Athersys RAGE
gb_hlc:AK004724	+	176.00	268.51	1.0e-05	2930	AK004724 Mus musculus adult ma
gb_est1:BG189682	+	175.00	279.72	2.5e-06	817	BG189682 RST8722 Athersys RAGE
gb_est1:BG208878	+	174.00	278.17	3.0e-06	808	BG208878 RST8269 Athersys RAGE
gb_est1:BG197314	+	173.00	276.57	3.7e-06	803	BG197314 RST16556 Athersys RAGE
gb_hlc:AK007863	+	172.00	267.33	1.2e-05	1704	AK007863 Mus musculus 10 day c
gb_est1:AM941785	+	171.00	272.34	6.4e-06	888	AM941785 GH05994.3p1me GH Dros
gb_gss:BG14276	+	170.50	274.35	4.9e-06	663	BG14276 2B8, Contig84 Subc
gb_gss:BG14466	+	170.50	274.35	4.9e-06	663	BG14466 2B8, Contig84 Subc
gb_est1:BG182853	+	170.00	271.23	7.3e-06	832	BG182853 RST17961 Athersys RAGE
gb_est1:BG189683	+	169.50	271.10	7.5e-06	776	BG189683 RST8722 Athersys RAGE
gb_gss:AB934336	+	169.00	273.07	5.8e-06	588	AB934336 BJ-Ba0002013 R. japo
gb_est1:BG110225	+	168.50	272.89	5.9e-06	551	BG110225 RST22387 Athersys RAGE
gb_est1:BG203018	+	167.00	266.38	1.3e-05	811	BG203018 RST19746 Athersys RAGE
gb_est1:BG200433	+	167.00	266.38	1.3e-05	821	BG200433 RST19220 Athersys RAGE
gb_est1:BG199924	+	166.50	265.62	1.5e-05	816	BG199924 RST19220 Athersys RAGE
gb_est1:BG133781	+	166.00	265.87	1.5e-05	733	BG133781 BJ133781 unpublished c
gb_hlc:AK002588	+	165.50	265.87	1.5e-05	708	AK002588 Mus musculus adult ma
gb_est1:BG202472	+	165.00	263.09	2.4e-05	824	BG202472 RST1829 Athersys RAGE
gb_est1:BG141013	+	164.00	262.03	2.4e-05	767	BG141013 RST1829 Athersys RAGE
gb_est1:BG206832	+	164.00	261.65	2.5e-05	790	BG206832 RST26134 unpublished c
gb_est1:BG194899	+	164.00	261.65	2.5e-05	801	BG194899 RST13956 Athersys RAGE
gb_est1:BI32937	+	163.50	261.19	2.7e-05	772	BI32937 602989330f1 NCI, CGAP-I
gb_est1:BI247661	+	163.00	263.26	2.0e-05	579	BI247661 602959155f1 NCI, CGAP-I
gb_est1:BI145034	+	163.00	259.82	3.2e-05	815	BI145034 602909101f1 NCI, CGAP-I
gb_est1:BG207705	+	162.00	259.13	3.5e-05	740	BG207705 RST727188 Athersys RAGE
gb_est1:BG191712	+	162.00	258.78	3.6e-05	766	BG191712 RST10808 Athersys RAGE
gb_est1:BI136721	+	160.00	256.59	4.8e-05	685	BI136721 unpublished c
gb_est1:BI150948	+	160.00	256.03	5.2e-05	724	BI150948 unpublished c
gb_est1:AL564426	-	158.00	250.63	0.0001	890	AL564426 LTI_NFY001_NH

```
gb_est2:BG185434 + 156.50 248.17 0.0001 887 | BG185434 RST4377 Athersys R
gb_est2:BG925017 + 156.00 249.89 0.0001 689 | BG925017 HNC5-1-F9 R HNC (H
gb_est2:BG200435 + 155.00 246.69 0.0002 802 | BG200435 RST19748 Athersys
gb_est2:BG205065 + 153.00 243.21 0.0003 815 | BG205065 RST19748 Athersys
gb_est1:AW257685 + 150.50 241.17 0.0003 661 | AW257685 SWD25CA00808SK B
```

```
seq_name: gb_hlc:AK004861
```

```
seq_documentation_block:
```

```
LOCUS AK004861 3004 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus adult male liver cDNA, RIKEN full-length enriched
library, clone:1300002013:cytoschrome P450, steroid inducible 3a11,
full insert sequence.
```

```
ACCESSION AK004861.1 GI:12836364
VERSION AK004861.1
KEYWORDS HTC; CAP trapper.
```

```
SOURCE Mus musculus (strain:C57BL/6J) adult male liver cDNA to mRNA,
clone:lib:RIKEN full-length enriched mouse cDNA library
clone:1300002013.
```

# ORGANISM

```
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

# REFERENCE

```
1 (sites)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
```

# JOURNAL

```
99279253
```

# MEDLINE

```
10349636
```

# PUBMED

```
2 (sites)
```

# REFERENCE

```
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
```

# JOURNAL

```
20499374
```

# MEDLINE

```
11042159
```

# PUBMED

```
3 (sites)
```

# REFERENCE

```
Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A., and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
```

# JOURNAL

```
20530913
```

# MEDLINE

```
11076861
```

# PUBMED

```
4 (sites)
```

# REFERENCE

```
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
```

# JOURNAL

```
5 (bases 1 to 3004)
```

# MEDLINE

```
11076861
```

# PUBMED

```
3 (sites)
```

# REFERENCE

```
Arakawa, T., Balderelli, R., Bono, H., Brownstein, M., Bull, C.,
Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T.,
Hara, A., Hayatsu, N., Hill, D., Hitamoto, K., Hiraoka, T., Horii, F.,
Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Katsukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koya, S.,
Kunihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, S.,
Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
Schriml, L., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T.,
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
Tanaka, T., Tejima, Y., Toyota, T., Yamamura, T., Yamakawa, I.,
Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and
Hayashizaki, Y.
```

Direct Submission  
Submitted (10-JUL-2000) Yoshihide Hayashizaki. The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),

RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216) Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGAGAGAGAGCGCCGCACTCGATGTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'-GAGAGAGAGAGAGATCAAGAGCTCAATTAATTAATTAACCCCGCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

## FEATURES

## Source

Location/Qualifiers  
1..3004  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="MGD:MGI:1907268"  
/db\_xref="taxon:10090"  
/clone="1300002013"  
/sex="male"  
/tissue.type="liver"  
/clone.lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
1..3004  
/gene="Cyp3a11"  
1..3004  
/gene="Cyp3a11"  
/note="cytochrome P450, steroid inducible 3a11  
data source:MGD, source key:MGI:88609, evidence:ISS"

BASE COUNT 836 a 701 c 702 g 765 t  
ORIGIN

## alignment\_scores:

Quality: 203.00 Length: 401  
Ratio: 1.063 Gaps: 13  
Percent similarity: 47.631 Percent identity: 22.195

## alignment\_block:

US-09-724-797-36 x AK004861 ..

Align seg 1/1 to: AK004861 from: 1 to: 3004

```
49 ValATGAlaValLeuGlyAspGluArgPheAlaArgThrGlyLeuArg 65
      ::::: ||||| ||| ||::: |||||
1297 ATTAGAGATGCTGCTAGAGGAATGTTTCTGCTTCACAAACCGCG 1346
      :::
65 gPheTrpThrAspLeuValGlyProGlyLeuLeuAlaGluLeuValGly 82
      ::: ||||| ::: ||:::
1347 GCAATTTT.....GGCCCA.....GTGGGATTAATGAGTA 1375
      ::: ||||| ::: ||:::
82 spLleLeuPheGlnAspGluProAspHisGlyArgLeuArgGlyVal 98
      ||| ::::: ||| ||| |||
1376 AAGCTATCTCAATCTTAAGGATGATGAGAGACATATAGACCTTGG 1425
      ::: ||::: |||
99 ValGlyProAlaPheSerProSerAlaLeuArgArgLeuGluProVal 115
      ::::: ||::: ||| ||| ||| |||
1426 CTGTCCCCACATTCACAGTGGAAACTCAAGGATGTTCCCTGTGAT 1475
      ::: ||::: |||
115 eAlaGlyThrValAspAspLeu.....LeuArgProAlaLeuAla 129
      ||| ||| |||
1476 TGAACAGTATGAGACATTTTGGTAAAGTACTTGAGCGCAGAAAGCA 1525
      ::: |||
129 TGGTAlaMetAspValValAspGluLeuAlaTyrProLeuAlaLeuArg 145
      ::||| |||
1526 AAGGCAAG.....CCTGTTACTATGAAA 1548
```

```
146 AlaValLeuGlyLeuLeuGlyLeuProAlaAlaAspTrpGlyAlaVal 162
      |||||
1549 GATGTGTA.....GG 1559
      :::
162 YArgTrpSerArgAspValGlyArgThrLeuAspArgGlyAlaSerAlaG 179
      ::||| ||||| ::: |||
1560 TGCTTACAGATGATGATGATGATCACCAGCATCATCTTTGGAGTAAC 1609
      ::: |||
179 LuAspMetArgArgGlyHisAlaAlaAlaGlu..... 190
      ::: |||
1610 ATTCCTCAACACACCAAGAGATCTTTGTGAGAAACCAAAAGCTT 1659
      ::: |||
190 ..... 190
1660 TTAAGATTGATTTTGTGATCCTTGTCTCTCTCAGTAGTACTTTTCC 1709
      ::: |||
190 ..... 190
1710 ATTCTGACACAGATATATGACATGTTAAATATCTGCATGTTCCAAAG 1759
      ::: |||
191 .....PheAlaAspTyrValGluArgAlaLeuAlaArgArg 202
      ||| ::||| |||
1760 ATTCATATAGATTTTTCAAAAATTTGTGACAGATGAAGAAACCGC 1809
      ::: |||
203 .....ArgArgGluGlyGlyAspLeuLeuAlaLeuMetLeuAs 216
      ::||| ||||| |||
1810 CTGATTCATTAAGCAGAAACCCAGTGGATTTCTTCACCTGATGTA 1859
      ::: |||
216 PAlaHisAspArgGly.....LeuMetSerArg 225
      ::||| |||
1860 TTCTCATTAATTAATTCACAAACAAAGTCTCTCATTAAGCCCTTTGAC 1909
      ::||| |||
226 AsnGluLeuValSerThrValValThrPheLeuPheThrGlyHisGlu 242
      ||||| ::||| ||||| |||
1910 ATGAGATATCAGCCAGCATATATCTTTATTTGGCGGATGAAC 1959
      ::: |||
242 rValAlaSerGlnValGlyAsnAlaValLeuSerLeuLeuAlaHisPro 258
      ::||| |||
1960 CACCAGTACACACATTTCTTCACCTGATTCCTTGGCCACACCTTG 2009
      ::: |||
259 .....AspGlnLeuAsp.....LeuLeuArgArg 266
      ::||| |||
2010 ATATCCAGAAAAACGTCAGAGATGATGATGATGAGCTCTGCCAACAG 2059
      ::||| |||
267 ArgPro.....AspLeuLeuAlaGlnAl 274
      |||
2060 GCACCTCCCACTATGATATCTGATGAGATGATGATGATGATGATGAT 2109
      ::||| |||
274 aValGluGlyCysLeuArgTyrAspProSerValGlnSerAsnThrArg 291
      ::||| ||| |||
2110 GCTTAATGAACCCCTCAGATTAATATCCATTCCTAATAGACTTGAGAG 2159
      ::: |||
291 InLeuAspValAspValGluLeuArgGlyArgArgLeuArgArgAsp 307
      ::||| |||
2160 TCTGTAAAGAAAGATGTTGAACTCAATGAGTGTATATCCCAAGGGTGA 2209
      ::: |||
308 ValValValValLeuAlaGlyAlaAlaAsnArgAspProArgTyrAs 324
      ::||| |||
2210 ACAGTATGATTCATCTTATGCTCTTCACACAGACACACAGACAGTGC 2259
      ::: |||
324 PArgProAspAspPheAspTlleGluArg..... 333
      ::||| |||
2260 AGAGCCGGAAGATTCACACTGAACCTGAGTTCACAGAGAACAGAGCA 2309
      ::: |||
334 .....AspProValProSerMetSerPheGlyAlaGlyMetArgTyr 348
      ::||| |||
2310 GCATTGATCTTATGTAATATCTGCCCTTTGGGATGAGACCAAGAACTGC 2359
      ::||| |||
349 LeuGlySerTyrLeuAlaArgThrGlnLeuArgAlaAlaValAlaAla 365
      ||||| ::||| |||
2360 CTTCGATGAGGTTTGTCTCATGAATATGAACCTGCTCTCAATAAT 2409
      ::||| |||
365 u 365
```

2410 T 2410

seq\_name: gb\_est1:AA986024

seq\_documentation\_block:

LOCUS	AA986024	768 bp	mRNA	linear	EST 28-MAY-1998
DEFINITION	uc72b04.y1 Sugano mouse liver mlia Mus musculus CDNA clone				

IMAGE:1431151\_5', similar to gb:X60452.M.musculus mRNA for cytochrome P-450I1A (MOUSE);, mRNA sequence.

ACCESSION	AA986024
VERSION	AA986024.1
	GI:3167413

KEYWORDS	EST.
SOURCE	house mouse.

ORGANISM *Mus musculus*  
Eukaryota; Metazoa; Ch

REFERENCE  
1 (bases 1 to 768)  
Mammalia; Eutheria; Rodentia; Muridae; Murinae; Murini; *Mus mus* L.

**AUTHORS**  
Marra, M., Hillier, L.,  
Geisel, S., Kucaba, T.,  
Schubert, R. Stettin

Schellenberg, K., Stept  
Theising, B., Wylie, T.,  
Intention D

waterston, R.  
The WashU-HHMI Mouse E  
ublished (1996)

**JOURNAL** Unpublished (1996)  
**COMMENT** Contact: Marra M/Mouse  
EASHT-HUNT MOUNO ECM T

Washnu-HHMI Mouse EST  
Washington University  
444 Forest Park Parkway

4444 Forest Park Parkway  
Tel: 314 286 1800  
Fax: 314 286 1810

306 GTGATGGAGATGGATAACCTGCATATGGTGCCTTAATGAACCCCTCAGATT 355

Source

```
1. .2492
/organism="Homo sapiens"
```

```
/db_xref="taxon:9606"
/clone="IMAGE:3354171"
/tissue_type="Eye, retinoblastoma"
/clone_lib="NIH MGC-16"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
BASE COUNT      512 a      727 c      679 g      574 t
ORIGIN

alignment_scores:
    Quality: 182.50      Length: 486
    Ratio: 0.882      Gaps: 21
    Percent similarity: 42.593      Percent identity: 23.045

alignment_block:
US-09-724-797-36 x BC001776 ..

Align seg 1/1 to: BC001776 from: 1 to: 2492

14 AlaAspValArgArgAspProTyrProSerTyr..... 24
|||||
166 GCAGACATCCAGAGCCCTCTACAGCCAGCTTCTGCGCCAACTTCTG 215
25 .....HistripleuLeuArgHisAspProValH 34
|||||
216 CAAGGGGGGCGTGTGAGCTACAGAGCTCAGGTGCGAGGGCGCGCCG 265
34 IsArgGlyAlaHisArgValTrrTyrValSerArgPheAlaAspValArg 50
|||||
266 ACTTCGGGCGG.....GTGTGGCTAGCCAGC...TTTGGAGCAGTGGC 306
51 AlaVal.....LeuGlyAspGluArgPheAlaArgTh 61
|||||
307 ACCGTGTACGTGTGCTGCCCCCTGCACCTGTGCGAGAGCTGCGACAGA 356
61 rGlyIleArg.....ArgPheTrrThrAsp..... 69
|||||
357 GGGACCCCGGCGCGAGCGCTCAGCTTCTCCCTGCGAGCGAGCACCGCC 406
70 .....LeuValGlyProGlyLeuLeuAlaGluIleValGly 81
|||||
407 GCTGCCCGCAGCGGCGCTTGGCGGA..... 429
82 AspIleIleuPheGlnAspGluProAspHisGlyArgLeuArgGlyVa 98
|||||
430 .....CTGCTCACTGCGGAGGGAGAGAAATGCAAGGCTCCCGCAGTCT 473
98 lValGlyPro...AlaPheSerProSerAlaLeuArgArgLeuGluProv 114
|||||
474 CTTGCCCGCGCTCCTCTCCGCGCTCAAGCGCGCGCTACGCGGAA 523
114 AlIleAlaGlyThrValAspAspleuLeuArg..... 124
|||||
524 CCCTGACACAGTACGTCTGCGAGCTTGTGCGGCGCTGTGAGCGCGCAG 573
125 .....ProAlaLeuAlaArgGlyAlaMetAspValVa 135
|||||
574 GGAGCTGGCAGCGGGCGCGCGCTGTGG.....GACGTGGC 614
135 lAspGluLeu..... 138
|||||
615 GGGGAATTTCACAAATTTCGAGCTGGAAGTGAATCCAGAGCAGAGCTG 664
139 .....AlaTyr 140
665 GGCAGGCGCTCGGGGCGCCTACCAAGACCTCCGGAACCTGACGGCGC 714
141 ProleuAlaLeuArgAlaValLeuGly.....LeuLeuGlyLe 153
|||||
715 CCCCTCCCGACAAAGCATCGCGCGGTTCGTGCTGCGCTCCGCGCTTGGGCT 764
153 uProAlaAlaAspTrrPglValAlaValGlyArgTrrSerArgAspValGly 170
```

```
|||||
765 GCGTGA.....GGCTCAAGTGGCACCCGACAGCGAGACCTTCATC 805
170 rGhrIleuAspArgGlyAlaSerAlaGluAspMetArgArgGlyHisAla 186
|||||
806 CCGCTGTGGGCTCGGTGTGTTCACAGCTTTGACATGCGATGCC 855
187 AlaIleAla..... 189
|||||
856 CCAGTGGCTGGCCACCTTGTGCTGGGCGCGCGCCCTCTGCCGAG 905
190 .....Glu.PheAlaAspTyrValGluArgAlaLeuAlaArgArg 202
|||||
906 ACTGGAGCAGATGTTGCA...TTTGTCAAGAGCAGCAGTGGAGCGGCGA 952
203 Arg.....ArgGluGlyGlyGluAspLeuAlaLeuMe 214
|||||
953 GAGCAGAGGCGACCCATGAGGAAGGAGGACAGCCCGAGAGAGCTTGA 1002
214 tLeuAspAlaHisAspArgGlyLeuMetSerArgAsnGlu..... 227
1003 GTCTGGGGCGCACCTGACCCACTTCTCTCCGGAGAGTTCCTGCC 1052
228 .....IleValSerThrValValThrPheIlePheThrGlyHisGluThr 242
|||||
1053 AGTCATCTCGGGAATGTGACAGAGTGTGCTATTGGCGGAGTGACAGC 1102
243 ValAlaSerIleValGlyAsnAlaValLeuSerLeuLeuAlaHisProAs 259
|||||
1103 GTGTCCAAACAGCTCTGTGGGCTGTGTATGAGCTCTCCCGGACCCGA 1152
259 P..... 259
1153 AGTCACAGACGACCTCCACTCAGAGATCACAGCTGCCCTGAGCGCTGCT 1202
260 .....GlnLeuAspLeuLeuArgArgArgProAspleuLeu 271
|||||
1203 CCAGTGGCTACCCCTCAGCCACTGTCTGTGCCAGCTGCC...CTGCTG 1249
272 AlAGlnAlaValGluGlyCysLeuArgTyrAspProSerValGlnSerAs 288
|||||
1250 AAGCGGTGTCAAGGAAGTGTAAAGACTGTACCTCTGTGCTAGTGA 1299
288 nThrArgGlnLeuAspValAspValGluLeuArgGlyArgArgLeuArgA 305
|||||
1300 TTCTCGTGTCCACAGACAAGACATTCATGTGGTGACTATATTATCCCA 1349
305 rGAspAspValValValIleuAlaGlyAlaAlaAsnArgAspProArg 321
|||||
1350 AAATACGCTGTGCTACTGTGTCACTATAGCCACTTCAAGGAGCCCTGCC 1399
322 ArgTyrAspArgProAspAspPheAspIleGluArg..... 333
|||||
1400 CAGTTCCACAGACCAATTTCTTCTCCAGCTCGCTGGCGGAGAGG 1449
334 .....AspProValProSerMetSerPheGlyAlaGlyMetArgT 347
1450 TCCACACCCCGACACCATTTGCATCTCTTCCCTTTGGCTTTGGCAGAGCA 1499
347 yrcysLeuGlySerTyrLeuAlaArgThrGlnLeuArgAlaAlaValAla 363
|||||
1500 GCTGTATGGGAGAGACCGCTGGCAGAGCTTGAATTCGAATTCGCTTTGGCC 1549
364 AlaLeu 365
1550 CAGATC 1555
seq_name: gb_hc:BC020267
seq_documentation_block: 2492 bp mRNA linear HTC 19-DEC-2001
LOCUS BC020267
DEFINITION Homo sapiens, clone IMAGE:4580963, mRNA.
ACCESSION BC020267
```

VERSION	BC020267.1	GI:17939623
KEYWORDS	HTC.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (bases 1 to 2492)	
TITLE	Straussberg,R.	
JOURNAL	Direct Submission	
	Submitted (19-DEC-2001)	National Institutes of Health, Mammalian

REMARK	NIH-MGC Project URL: <a href="http://mgc.ncl.nih.gov">http://mgc.ncl.nih.gov</a>
COMMENT	Contact: MGC help desk

BC Cancer Agency, Vancouver, BC, Canada  
info@bccsc.bc.ca

Steven Jones, Jennifer Aseno, Ian Bosdel, Aaron Butterfield, Susanna Chan, Readman Chin, Chris Felli, Erin Garland, Ran Guin, Letticia Hsiao, Martin Krywaniuk, Reta Kutscher, Oliver Lee, Soo Seon Lee, Victor Ling, Carrie Mathewson, Candice McLeay, Steven Ness, Pawan Pandoh, Anna-Jill Prabhu, Parvaneh Speedi, Jacqueline Schein, Duane Smallus, Michael Smith, Lorraine Spence, Jeff Stett, Michael Thorne, Miranda Tsai, Natasha van den Bosch, Jill Vardy, George Yonge, Scott Zuyderduyn, Marco Marra.

clone distribution: MCC clone distribution information can be found through the I.M.A.G.E. Consortium/INL at: <http://image.lnl.gov>  
Series: IRAT Plate: 40 Row: e Column: 19  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 13904864  
This clone has the following problem: no cloning site / microdeletion.

FEATURES	Location/Qualifiers
source	1. .2492

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4580963"
/rnaise_type="Eye, retinoblastoma"
/clone_lib="N1H_MGC_16"
/adb_hsc="DH10B-R"
/notes="vector: pOTB7"

```

```
alignment_scores:
  quality: 182.50
  ratio: 0.882
  percent_similarity: 42.593
  percent_identity: 23.045
```

alignment\_block:

Align seg 1/1 to: BC020267 from: 1 to: 2492

14 AlaaspValAlaArgAspProtyrProSerIyr ..... 24  
 166 GCAGACATCCCAAGCCCTCTACGCCAGCTTTCTGGCCGAATTTTCTG 215  
 25 ..... HistProLeuAlaArgHisAspProValH 34  
 216 CAAGGGGGGCGTCGAGGCTACACGAGCTCTCAGTGCAGGGCCGCCGCG 265  
 34 IsArgIyAlaHAlaHisArgValTPrTyIyLserArgPheAlaAspValArg 50  
 266 ACTTTCGGGCGG.....GTGGGCTAGGCCAAC...TTTGGGACAGTGC 306

51 AlaVala.....LeuGlyAspGluArgPheAlaArgTh 61  
52 AspiLeileLeuPheGlnAspGluProaspHisGlyArgLeuArgGlyVala 98  
53 ACCGGTACGTCGGCTGCCCTTCGACTGCTGCAGAGAGCTGCTGCACAGAGA 356  
54 rGlyIleArg.....ArgPheTrpHisAsp..... 69  
55 ||| ||| |||||  
56 GGGACCCCGCGCCGAGCGCTGACGTTCTCCCTCGGAGGAGGACCGCC 406  
57 .....LeuValGlyProGlyLeuLeuAlaGluIleValGly 81  
58 |||  
59 GCTGCCGCCACCGGGGCTTGCAGA..... 429  
60  
61 AspiLeileLeuPheGlnAspGluProaspHisGlyArgLeuArgGlyVala 98  
62 .....CTGCTACACTGCGGAGAGGAGGAGAAAGGCAAGGCTCCGACAGCT 473  
63 |ValGlyPro..AlaPheSerProSerAlaLeuArgArgLeuGluProVal 114  
64 CCGGCCCCCGCTCTCTCTCCCGGCTCAAGCGCGCCCGCTACGCGCGAGA 523  
65 |||||  
66 alIleAlaGlyThrValAspAspLeuArg..... 124  
67 |||||  
68 CCGTACACACAGCTAGTCTGCACACTTGGCGGCGCTGAGGCGCCACCGG 573  
69 .....ProAlaLeuAlaGlyAlaMetAspVala 135  
70 GACGTGCGACGCGGGCGCCCGCCCTGCTGG.....GACGTGCG 614  
71 LaspGluLeu..... 138  
72 |||||  
73 GGGGGAATTTTACAAATTGCGACTGGAAGTGAGTCCAGACAGACTG 664  
74 .....AlaTyr 140  
75 GGCAGCGTACGAGGGGCGGCTTACCAGAGCTCCCGAACCCTGACGGCGC 714  
76 ProLeuAlaLeuArgAlaValLeuGly.....LeuLeuGlyLe 153  
77 ||||| ||| |||||  
78 CCCCTCCCGACAGGCAATGCCCGCGTCTGCTGCGCTGCGCTGGCT 764  
79 uProAlaIlaAspTrpGlyAlaValGlyArgTrpSerArgAspValGly 170  
80 ||||| |||||  
81 GCGCTGTGGCTCGTGTGTGTGTGTCTCACGCTGTGACCAATGCGAGTCC 855  
82 GCGTGA.....GGCTCAAGTGCACCCGACCGAGAGACCTTCATC 805  
83 AlaIleAla..... 189  
84 |||||  
85 CCACGTGCTGGCGCCACCTTGGCTCGGGCCCTGGGGCCGCTCTGCCAG 905  
86 .....Glu.PheAlaAspTyrValGluArgAlaLeuAlaArgArg 202  
87 ACTGGAGCCAAATGTTTGA...TTTGCTCAGAGGCACTGAGACGGCGGA 952  
88 Arg.....ArgGluGlyGlyGlnAspLeuLeuAlaIle 214  
89 |||||  
90 GAGGCGAGGACCAACCATGAGACGAGAGACACGCCCGAAGACACTGGA 1002  
91 LeuAspAlaHisAspArgGlyLeuMetSerArgAsnGlu..... 227  
92 |||||  
93 GTCGTGGCGGACCTCGACCACTTCCTCTTCCGGGAGAGATGGCTGCC 1052  
94 .....IleValSerThrValValThrPheIleIlePheThrGlyHisGluThr 242  
95 ||||| |||||  
96 AGTCATCTCTGGGAATGTGACAGAGTTGATTTGGCGGAGTGCACAGC 1102  
97 ValAlaSerGlnValGlyAsnAlaValLeuSerLeuLeuAlaHisProAs 259  
98 GTGTCAACACAGCTCTCTTGGGCTCTGTATAGCTCTCCGGGACCCCGA 1152  
99 P..... 259

```

1153 AGTCAGACAGACGCTCCACTCAGATCAGAGCTGCCTGAGCCCTGGCT 1202
260 .....GlnLeuAspLeuLeuArgArgProAspLeu 271
1203 CCAGTCCCTACCCCTCAGCCACTGTCTGTCCAGCTGCC...CTGCTG 1249
272 AlaGlnAlaValGluGluCysLeuArgTyrAspProSerValGlnSerAs 288
1250 AAGCGGTGTGTCAGAGAGTGTAGACTGTACCTGTGTGTAACCTGAGAA 1299
288 ntrHArgGlnLeuAspValAspValGluLeuArgGlyArgArgLeuArg 305
1300 TTCTGTGTCCAGACAAAGACATTCATGTGGTACTATATATATATATAT 1349
305 rGAspAspValValValAlaValAlaGlyAlaAlaAsnArgAspProArg 321
1350 AAAATAGCGTGTGCTACTGTGTGCTACTATGCACTTCAAGGACCTTGCC 1399
322 ArgTyrAspArgProAspAspAspIleGluArg..... 333
1400 CAGTTCCTCCAGAGCCAAATCTTTCTCCAGCTCGCTGGGGAGAGG 1449
334 .....AspProValProSerMetSerPheGlyAlaGlyMetArgT 347
1450 TCCACACCCCAACCCATTGTCATCTCTCTTGGCTTGGCAGAGCGCA 1499
347 yrcysLeuGlySerTyrLeuAlaArgTyrGlnLeuArgAlaAlaValAla 363
1500 GCTGTGTGGGAGAGCCCTGCGCAGAGCTGTAATGCAATGGCTTTGGCC 1549
364 AlaLeu 365
1550 CAGATC 1555

seq_name: gb_est2:BI246899
seq_documentation_block:
LOCUS BI246899 717 bp mRNA linear EST 17-JUL-2001
DEFINITION 602959455F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5125068 5',
ACCESSION BI246899
VERSION BI246899.1 GI:14791308
KEYWORDS mRNA sequence.
SOURCE EST.
ORGANISM house mouse.
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 717)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rgs@bbs-rcmail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L1AM1305 row: n column: 13
High quality sequence stop: 717.
Location/Qualifiers
1..717
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5125068"
/clone_lib="NCI CGAP L19"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT."

```

```

Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 212 a 175 c 150 g 180 t
ORIGIN
alignment_scores:
    quality: 182.00    length: 179
    ratio: 1.784      gaps: 5
    percent similarity: 56.983    percent identity: 29.050
alignment_block:
US-09-724-797-36 x BI246899 ..
Align seg 1/1 to: BI246899 from: 1 to: 717
209 AspleuLeuAlaLeuMetLeuAspAlaHisArgGly..... 221
|||||  |||||:|||||:|||||:|||||:|||||:
61 GATTCTCTCAGCTGATGATGATGATCTCATATATATCCAAAGCAAGT 110
222 .....LeuMetSerArgAsnGluIleValSerThrValAlaIhrp 235
|||||  |||||:|||||:|||||:|||||:|||||:
111 CTCTCATAAAGCCCTTCTGACATGAGATCACAGCCAGTCATATATCT 160
235 heilePheThrGlyHisGluThrValAlaSerGlnValGlyAsnAlaVal 251
|||||  |||||:|||||:|||||:|||||:|||||:
161 TTATTTTGTGCTGGTATGAACACACAGTACACTTCTCCACCTG 210
252 LeuSerLeuLeuAlaHisPro.....AspGlnLe 261
|||||  |||||:|||||:|||||:|||||:
211 CATTCCTTGGCCACTCACCTGATATCCAGAAAAAATGACAGATGAT 260
261 uAsp.....LeuLeuArgArgArgPro..... 268
|||||  |||||:|||||:|||||:|||||:
261 CGATGAGGCTCTGCCACACAGGACCTCCACGATGATCTGATGAGG 310
269 .....AspleuLeuAlaGlnAlaValGluCysLeuArgTyrAspPro 283
|||  |||:|||||:|||||:|||||:
311 AGATGGAATACCTGATGATGATGATGATGATGATGATGATGATGATGAT 360
284 SerValGlnSerAsnThrArgGlnLeuAspValAspValGluLeuArgG 300
|||||  |||||:|||||:|||||:|||||:
361 ATGGCTAATGACTTGAAGAGCTGTGAAGAAAGTGTGAACCTCAATGG 410
300 yArgArgLeuArgArgAspAspValValValAlaValAlaAlaAla 317
|||  |||:|||||:|||||:|||||:
411 TGTGTATATCCCAAGGGTCAACAGTATGATTCATCTTATGCTTTC 460
317 snArgAspProArgArgTyrAspArgProAspAspAspIleGluArg 333
|||||  |||||:|||||:|||||:|||||:
461 ACCATGACCCACAGACGCTGACAGCTGAAGAAATCCAACTTCAAGG 510
334 .....AspProValProSerMetSerPhe 341
|||||  |||||:|||||:|||||:
511 TTGAGCAAGAGAAACAAGGACATTCCTTATGATATCTCCCTT 560
341 eGlyAlaGlyMetArgTyrCysLeuGlySerTyrLeu 353
|||||  |||||:|||||:|||||:|||||:
561 TGGGAATGGACCCAGGAACCTGTGCTGCAATGAGGTTG 597

seq_name: gb_est2:BG204536
seq_documentation_block:
LOCUS BG204536 789 bp mRNA linear EST 21-APR-2001
DEFINITION R8723942 Atheros RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG204536
VERSION BG204536.1 GI:13726223
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 789)

```

```

292 euaspValaspValagluLeuarGlyArgArgrLueurArgAspPval 308  

    ||| ||||| :::: :  

460 GCAAAAAGATGTGGATCATTCATGGGATGTTCTCCAAAGGGCGTGT 509  

309 ValValalLeualacLYlaLaalnargspProrAarAgLyrrasAp 325  

    ||| :::: :  

510 GTGATGATGCCAGCATACCTCTCACCGTGACCACCAAGTACTGCACAGA 559  

325 gPriAspPhesppHlegLuArg..... 333  

    ||| :::: :  

560 GCCGGAACAAGTCTCCTCGTAAGAAAGATTGACAAGAAGAACAAGAACAA 609  

334 ..AsprvAlProSerMetserPheglyAlagLYmeLrTyCyLau 349  

    ||| ||||| ||||| :::  

610 TAGATCTTCATTACATACACCCTTGAGAGTGAGCCACCAAACTGCATY 659  

350 GlyseerTYreuaLaalAgrThrgInleuarAgalaLaValaLaLeual 366  

    ||| :::: :  

660 GGCAITGAGTTTGCTCTCATGAACATGANA.....CTTGCTCPAT 700  

366 AARGLEUPR 369  

    ||| |||  

701 CAGAGTCTT 710  

seq_name: gb_est2.BG206682  
  

seg_documentation_block: 755 bp mRNA linear EST 21-Apr-2001  

LOCUS BG206682 RST26133 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.  

DEFINITION BGI206682  

ACCESSION BG206682 GI:13728369  

VERSION BG206682.1 GI:13728369  

KEYWORDS EST.  

SOURCE human.  

ORGANISM Homo sapiens  

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  

REFERENCE 1 (bases 1 to 755)  

AUTHORS Harrington,J.,J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,  

Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J,  

'Lerner,L.', Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith  

'E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher  

J., Danzig,J. and Ducat,M.  

Creation of genome-wide protein expression libraries using random  

activation of gene expression  

Nat. Biotechnol. 19 (5), 440-445 (2001)  

COMMENT Contact: Scott J. Cain  

Athensys, Inc.  

3201 Carnegie Ave.  

Tel.: 216 431 9900  

Fax: 216 361 9596  

Email: scaine@athensys.com  

High quality sequence stop: 496.  

FEATURES  

source location/Qualifiers  

1..755  

/organism="Homo sapiens"  

/db_xref="taxon:9606"  

/cloae_lib="Athensys RAGE Library"  

/cell_line="HT1080"  

/note="See 'Creation of Genome-wide Protein Expression  

Libraries using Random Activation of Gene Expression',  

Nature Biotechnology, In press. Note that even though the  

cell type indicated is HT1080, such as a random activation  

method was used, these sequence tags are not necessarily  

expressed in HT1080 under normal circumstances."  

BASE COUNT 231 a 171 c 153 g 200 t  

ORIGIN  

alignment_scores: Quality: 179.00 Length: 213  

Ratio: 1.584 Gaps: 4
```



Percent Similarity: 53.052 Percent Identity: 25.822

alignment\_block:

US-09-724-797-36 x BG206682 ..

Align seg 1/1 to: BG206682 from: 1 to: 755

```

190 GIuphealaasptyrValGluarGalaLeuAlaArgArgArgGluG1 206
    ||| ..... ||| .....
60 GAAGTTACAAATTTTAAAGAAATCTGTAAAAAGATGAAGAAGACGC 109
206 yglYglu..... AspleuLeuAlaLeuMetLeuA 216
    ||| ..... ||| .....
110 CCTCGAAGATACACAAAGACCGAGTGTGATTCCTTCAGCTGATGATG 159
216 spAlaHisasp..... ArgGlyLeuMetSerArgasn 226
    ||| ..... ||| .....
160 ACTCTCAAAATTCAAAAGAACTGAGTCCCAAAAGCTCTCTCCGATCTG 209
227 GIuileValSerThrValValThrPheIlePheThrGlyHisGluThrVa 243
    ||| ..... ||| .....
210 GAGTCGCGGCCCAATCAATATCTTTATTTTGTGCTGCTATGAACCCAC 259
243 lAlaSerGlnValGlyAsnAlaValLeuSerLeuLeuAlaHisProaspG 260
    ||| ..... ||| .....
260 GAGCAGTGTCTCTCTTCATTATGATGATGAGACTGCGCACCCTCATG 309
260 lInleuAspleuLeuArgArgArgProasp..... 269
    ||| ..... ||| .....
310 TCCAGCAAAACTGACAGAGGAATGTGATGCACTTTTACCAATAAGGCA 359
270 ..... LeuLeuAlaGlnAlaVa 275
    ||| ..... ||| .....
360 CCACCCACTATGATAGTGTCTGACAGATGAGATCTTGACATGATGGT 409
275 lGluGluCysLeuArgArgProSerValGlnSerAsnThrArgGlnL 292
    ||| ..... ||| .....
410 GAATGAACGCTCAGATTAATCCCAATGTGATGAGACTGAGAGGGTCT 459
292 euAspValAspValGluLeuArgGlyArgLeuArgArgArgAspVal 308
    ||| ..... ||| .....
460 GCAAAAAAGATGTGAGATCAATGAGATGTTCATCCCAAGGGGTGTG 509
309 ValValValLeuAlaGlyAlaAlaAsnArgAspProArgArgTyAspAr 325
    ||| ..... ||| .....
510 GTGATGATGCCAGCTATGCTCTTACCGCCCAAGTACTGACAGAGA 559
325 gProAspAspPheAspIleGluArg..... 333
    ||| ..... ||| .....
560 GCCTGAGAAAGTCTCCCTGAAAGATTCACAGACAGACAGACAGACA 609
334 ..AspProValProSerMetSerPheGlyAlaGlyMetArgTyCyLeu 349
    ||| ..... ||| .....
610 TAGATCTCTACATATACACACACCTTTGGAGATGAGACCCAGAACTG 659
350 GlySerTyLeuAlaArgThrGlnLeuAlaAlaVal 362
    ||| ..... ||| .....
660 GGCATGAGGTTTGTCTCATGACATGATAAACTTGCTCTA 698

```

seq\_name: gb\_est2:BG182832

seg\_documentation\_block:

LOCUS BG182832 765 bp mRNA linear EST 21-APR-2001  
 DEFINITION RST1693 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BG182832  
 VERSION BG182832.1 GI:13704503  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 765)  
 AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,

Cain,S., Leventhal,C., Thornton,M., Ramchandran,R., Whittington,J.,  
 Lerner,L., Costanzo,D., McElligott,K., Boozet,S., Mays,R., Smith  
 E., Veloso,N., Kika,A., Hess,J., Colhoun,K., Lo,K., Offenbacher  
 J., Danzig,J. and Ducar,M.  
 Creation of genome-wide protein expression libraries using random  
 activation of gene expression  
 Nat. Biotechnol. 19 (5), 440-445 (2001)  
 21227151

COMMENT Contact: Scott J. Cain

Athersys, Inc.  
 3201 Carnegie Ave, Cleveland, OH 44115, USA  
 Tel: 216 431 9900  
 Fax: 216 361 9596  
 Email: scain@atersys.com

FEATURES High quality sequence stop: 506.

source Location/Qualifiers

1..765  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Athersys RAGE Library"  
 /cell\_line="HT1080"

/note="See 'Creation of Genome-wide Protein Expression  
 Libraries using Random Activation of Gene Expression',  
 Nature Biotechnology, in press. Note that even though the  
 cell type indicated is HT1080, since a random activation  
 method was used, these sequence tags are not necessarily  
 expressed in HT1080 under normal circumstances."

BASE COUNT 237 a 172 c 154 g 202 t

ORIGIN

alignment\_scores: Quality: 179.00 Length: 213  
 Ratio: 1.584 Gaps: 4  
 Percent Similarity: 53.052 Percent Identity: 25.822

alignment\_block:

US-09-724-797-36 x BG182832 ..

Align seg 1/1 to: BG182832 from: 1 to: 765

```

190 GIuphealaasptyrValGluarGalaLeuAlaArgArgArgGluG1 206
    ||| ..... ||| .....
60 GAAGTTACAAATTTTAAAGAAATCTGTAAAAAGATGAAGAAGACGC 109
206 yglYglu..... AspleuLeuAlaLeuMetLeuA 216
    ||| ..... ||| .....
110 CCTCGAAGATACACAAAGACCGAGTGTGATTCCTTCAGCTGATGATG 159
216 spAlaHisasp..... ArgGlyLeuMetSerArgasn 226
    ||| ..... ||| .....
160 ACTCTCAAAATTCAAAAGAACTGAGTCCCAAAAGCTCTCTCCGATCTG 209
227 GIuileValSerThrValValThrPheIlePheThrGlyHisGluThrVa 243
    ||| ..... ||| .....
210 GAGCTCTGTGCCCCAATCAATTAATCTTTTGTGCTGCTATGAACCCAC 259
243 lAlaSerGlnValGlyAsnAlaValLeuSerLeuLeuAlaHisProaspG 260
    ||| ..... ||| .....
260 GACAGTGTCTCTCTCTCATTAATGATGATGAGCTGACCTCAGATG 309
260 lInleuAspleuLeuArgArgArgProasp..... 269
    ||| ..... ||| .....
310 TCCAGCAAAACTGACAGAGGAATGTGATGCACTTTTACCAATAAGGCA 359
270 ..... LeuLeuAlaGlnAlaVa 275
    ||| ..... ||| .....
360 CCACCCACTATGATAGTGTCTGACAGATGAGATCTTGACATGATGGT 409
275 lGluGluCysLeuArgArgProSerValGlnSerAsnThrArgGlnL 292
    ||| ..... ||| .....
410 GAATGAACGCTCAGATTAATCCCAATGTGATGAGACTGAGAGGGTCT 459

```

292 euaspValaspValgluLeuAargLysArgArgLeuAargArgaspVal 308  
|||||.....  
460 GCAAAAAAGATGTTGAGATCAATGGATGTTCCCAAGGGGTG 509  
309 ValValValLeuAagLysAlaAlaAsnArgAspProArgArgTyrAsp 325  
|||||.....  
510 GTATGATTCACAGCTATGCTCTTACACCGTACCAAGACTGACAG 559  
325 gProAspPheAspIleGluArg..... 333  
:|||||.....  
560 GCCTGAGAAAGTCTCCCTCGAAGATTACGAAAGAACAGACAGACA 609  
334 . AspproValProserMetSerPheGlyAlaGlyMetArgTyrCysLeu 349  
|||||.....  
610 TAGATCTCTTCATATACACACCTTTGGAGTGGACCCAGAACTGATT 659  
350 GlySerTyrLeuAlaArgThrGlnLeuArgAlaVal 362  
|||||.....  
660 GGCATGAGGTTTGCTCATGAAACATGAACCTTGCTCTA 698  
seq\_name: gb\_est1:AI323114  
seq\_documentation\_block: 814 bp mRNA linear EST 23-DEC-1998  
LOCUS AI323114  
DEFINITION mJ87b03.v1 Soares mouse p3NMF19.5 Mus musculus cDNA clone  
IMAGE:483053.5, similar to gb:x63023.M.musculus mRNA for cytochrome  
P-45011A (MOUSE);, mRNA sequence.  
ACCESSION AI323114  
VERSION AI323114.1 GI:4057543  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 814)  
Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HMT Mouse EST Project  
Unpublished (1996)  
Contact: Maria M/Mouse EST Project  
WashU-HMT Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MG1:293797  
This read is a RESEQUENCE of a previously sequenced mouse clone  
This read has been verified (found to hit its original self in the  
correct orientation)  
Seq primer: -40RP from Gibco  
High quality sequence stop: 450.  
Location/Qualifiers  
1..814  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone\_lib="Soares mouse p3NMF19.5"  
/dev\_stage="19.5 dpc total fetus"  
/note="Vector: pT73D (ampicillin resistant)"  
/note="Vector: pT73D (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'  
TGTACCAATCGAAGTGGAGCGGCGGATATTTTATTTTATTTT 3']  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library went through one round of

normalization to a Cot = 5. Library constructed by Bento  
Soares and M.Fatima Bonaldo. RNA was kindly provided by  
Dr. Minoru Ko (Wayne State University)."  
BASE COUNT 252 a 167 c 182 g 211 t 2 others  
ORIGIN  
alignment\_scores: Quality: 177.50 Length: 187  
Ratio: 1.723 Gaps: 5  
Percent Similarity: 55.080 Percent Identity: 28.342  
alignment\_block:  
US-09-724-797-36 x AI323114 ..  
Align seg 1/1 to: AI323114 from: 1 to: 814  
209 AspleuLeuAlaLeuMetLeuAspAlaHisasp..... 219  
|||||.....  
38 GATTTTCTTCACGCTGATGATTAACCTCCAGAAATTACAAACCAAGAAATC 87  
220 .ArgGlyLeuMetSerArgAsnGlnIleValSerThrValValThrPheI 236  
:|||||.....  
88 TCATTAAGCATTATCTGATGTGGAGATTGTGGCTCAGTCAGTATCTTTA 137  
236 lApheThrGlyHisGlnThrValAlaSerGlnValGlyAsnAlaValLeu 252  
|||||.....  
138 TTTTGTGCTGCTATGAGACACACAGAGTCTTCTTGTGATTTGAT 187  
253 SerLeuLeuAlaHisPro.....AspGlnLeuAs 262  
|||||.....  
188 TTGCTGGCCATTACACCTGATGTACAGAAAGAACTTCAGATGAATGA 237  
262 P.....LeuLeuArgArgArgPro..... 268  
238 TCGGCTGTCGCCAATAGGACCTGCCACTATATACCTGCTACGA 287  
269 .AspleuLeuAlaGlnAlaValGluGlyCysLeuArgTyrAspProser 284  
:|||||.....  
288 TGGAGATATTCAGATCATGTGTGTAATGAACCTCAGATTATATCCATTC 337  
285 ValGlnSerAsnThrArgGlnLeuAspValAlaValGluLeuArgGlyAr 301  
|||||.....  
338 GCTGGAAGACTTGAGAGGCTCTGTAAAGACAGATTTGAAATCAAGGGCT 387  
301 gArgLeuAargArgAspAspValValValLeuAlaGlyAlaAlaAsnA 318  
:|||||.....  
388 ATTCATTTCCCAAGGAGCTGTGTGTATACCAACCTTGCTCTTCACA 437  
318 rGAspProArgArgTyrAspArgProAspAspIleGluArg... 333  
:|||||.....  
438 AAGACCCGAGATGACGTGGCCAGAGCTGAGGAATTCGCCCTGAAGGTTTC 487  
334 .....AspproValProserMetSerPheG1 342  
:|||||.....  
488 AGCAAGAAAGATCANGACAGCATCAATCCTTACATGTACTCCCTTGG 537  
342 yAlaGlyMetArgTyrCysLeuGlySerTyrLeuAlaArgThrGlnLeuA 359  
:|||||.....  
538 GAGTGAACCAAGGAACTGCTTGGCATGAGGTTTGTCTCATAAACATGA 587  
359 rGAlaAlaVal 362  
:|||||.....  
588 AAGTGTCTCTT 598  
seq\_name: gb\_est2:BG195942  
seq\_documentation\_block: 828 bp mRNA linear EST 21-APR-2001  
LOCUS BG195942  
DEFINITION R815025 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.  
ACCESSION BG195942  
VERSION BG195942.1 GI:13717517  
KEYWORDS EST.

```

SOURCE          human.
ORGANISM        Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE       1 (bases 1 to 828)
AUTHORS         Harrington,J.J., Sherif,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Leventhal,C., Thornton,D., Ramachandran,R., Whittington,J.,
Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith
,E., Veloso,N., Klika,A., Hess,J., Cotlren,K., Lo,K., Offenbacher
,J., Danzig,J. and Ducar,M.
TITLE           Activation of genome-wide protein expression libraries using random
JOURNAL         Nat. Biotechnol. 19 (5), 440-445 (2001)
MEDLINE         21227151
COMMENT         Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 549.
FEATURES
Source          location/Qualifiers
1..828
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression
libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances.."
BASE COUNT      249 a 191 c 165 g 220 t 3 others
ORIGIN
alignment_scores:
Quality: 176.50 Length: 248
Ratio: 1.379 Gaps: 5
Percent Similarity: 51.613 Percent Identity: 24.597
alignment_block:
US-09-724-797-36 x BG195942 ..
Align seg 1/1 to: BG195942 from: 1 to: 828
190 GUGPheAlaSPYrValGluArgAlaLeuAlaArgArgArgGluG 206
||| :::::::::::::::::::: ||| :::::::::::
96 GAAGTACAAATTTTAAAGAAATCTGTAAAGATGAAGAAGAGTCG 145
206 yGlyGlu.....AspLeuAlaLeuMetLeuA 216
||| ::::::::::||| :::::::::::
146 CCTCGAAGATTACAAAGACACGAGTGGATTCTTCACCTGATGATG 195
216 sPAlaHisAsp.....ArgGlyLeuMetSerArgAsn 226
||| :::::::::::::::::::: ||| :::::::::::
196 ACTCTGAGATTCAAAAGAACTGAGTCCCAATGCTTGTCCGATCTG 245
246 GAGCTGCTGCGCCCAATCAATTATCTTTATTTTCTGCTGATGAACAC 295
243 lAlaSerGlnValGlyAsnAlaValLeuSerLeuLeuAlaHisProAspG 260
:::||||| ::::::::::: ||| :::::::::::
296 GACCAATGTTCTCTCCCTCATTTATATGAAGCTGACCTACCCGTATG 345
260 lIneuAspLeuArgArgArgProAsp.....269
::: ||| ::::::::::: |||
346 TCCAGCAGAAACTGCAGAGAAATTGATGACGTTTAACCCAATAAAGCA 395
270 .....LeuLeuAlaGlnAlaVala 275
||| ::::::::::|||

```

396 CCACCACCTATGATAGTGTGCTACAGATGGAGATATTGACATGGTGTG 445

275 TGIUGUCysLeuArgTyrAspProSerValGlnSerAsnThrArgGln 292

446 GAATGAAACCGCTCAGATTATTCCTCAATTGGTATGGAGCTGAGGGCTCT 495

292 euAspValAspValGlnLeuArgGlyArgArgLeuArgArgAspAspVal 308

496 GCAAAAAAGATGTTGAGATCAATGGAGATGTTTCATCCCAAAAGGGGTGGT 545

309 ValValValLeuValGlyAlaAlaAsnArgAspProArgTyrAspArg 325

546 GTGATGATCCCAAGCTATGCTCTTCACCGTGCACCCANAGTACTGGACAGA 595

325 gProAspAspPheAspIleGluArg..... 333

596 GCGTTAGAAAGTTCTCCTCGTAGAAGATTGACCAAGAAACAAGACACACA 645

334 ..AspProValProSerMetSerPheGlyAlaGlyMetArgTyrCysLeu 349

646 TAGATCCTTACATATACACACCCCTTTGGAGAGTGAGCCACCAACATCGAT 695

350 GlySerTyrLeuAlaArgThrGlnLeuArgAlaAlaValAlaAlaLeuAl 366

696 GCGTTGAGGTTGCTCTCATGACATGAAACCTGCTCTATATAGAGTCTT 745

366 argLeuProGlyLeuArgGlyCysAlaSerAspAlaLeuAlaTyrG 383

746 TGAGAACTTCCTCTTAACCTTGTAAAGAACACAGAT..... 782

383 InProArgThrMetPheArgGlyLeuAlaSerLeuProIleAla 397

783 ..CCGCTGAGCTCTATAAGAGATCAACATCTTCATATTTC 824

seq\_name: gb\_hnc:AK004724

seq\_documentation\_block:

LOCUS AK004724 2930 bp mRNA linear HTC 19-JAN-2002

DEFINITION Mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:1200012E24;Kallikrein B, plasma 1, full insert sequence.

ACCESSION AK004724

VERSION AK004724.1 GI:12836110

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (Strain: C57BL/6J) adult male lung cDNA to mRNA, clone:1200012E24.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (sites)

AUTHORS Carninci,P. and Hayashizaki,Y.

TITLE High efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 99279253

PUBMED 10349636

REFERENCE 2 (sites)

AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE 20499374

PUBMED 11042159

REFERENCE 3 (sites)

AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P., Kono,H., Akiyama,Y., Nishik,S., Kitsuana,T., Iashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishimi,T., Harada,A., Yamamoto,S., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kaishiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

TITLE RIKEN integrated sequence analysis (RISA) system--384-format





[illegible]



